

全国高等职业教育应用型规划教材·经济管理基础课

统计学基础

(第4版)

阮红伟 主 编

康燕燕 副主编

李莉莉 主 审

電子工業出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

《统计学基础》(第4版)是在第3版的基础上修改而成的,在保留第3版内容框架的基础上,删除了一些陈旧的内容,增进了统计在目前社会经济生活中的应用,同时更加突出统计的实用性和时代性。本书是一本统计入门读物,共有9章内容,包括总论、统计调查、统计整理、总量指标和相对指标、平均指标和标志变异指标、抽样推断、时间数列、统计指数及相关分析与回归分析。

本书结构模块包括学习要点、正文、统计术语、重点知识梳理、习题与实践训练、统计学应用案例。本书以统计工作流程为主线,始终贯彻学以致用、理论联系实际的原则,注重实践能力和创新精神培养,理论适中、案例丰富、操作性强,具有鲜明的时代性和较强的实用性。

本书既可以作为高等院校经济类、管理类各专业的教学用书,也可供社会培训班使用或作为自学用书,对广大的实际工作者也极具参考价值。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

统计学基础 / 阮红伟主编. —4版. —北京:电子工业出版社, 2016.1

ISBN 978-7-121-27689-7

I. ①统… II. ①阮… III. ①统计学—高等教育—教材 IV. ①C8

中国版本图书馆 CIP 数据核字(2015)第 284162 号

策划编辑:贾瑞敏

责任编辑:贾瑞敏

特约编辑:田学清 赵海红

印 刷:

装 订:

出版发行:电子工业出版社

北京市海淀区万寿路 173 信箱

邮编:100036

开 本:787×1092 1/16

印张:20.75

字数:564千字

版 次:2005年8月第1版

2016年1月第4版

印 次:2016年1月第1次印刷

印 数:4000册 定价:43.50元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888。

质量投诉请发邮件至 zlt@phei.com.cn, 盗版侵权举报请发邮件到 dbqq@phei.com.cn。

服务热线:(010)88258888。

前 言

《统计学基础》(第4版)是在我们二十多年来对经济类、管理类各专业学生统计学教学实践经验的基础上,经总结修改而写成的。本教材自2005年8月第1版、2008年9月第2版、2012年7月第3版以来,均先后印刷多次,获得广大学生和读者的接受和肯定。本次第4版进行修订之前,我们通过问卷调查及座谈方式,对部分教师和读者进行了调查,在广泛吸取读者意见的基础上,对第3版中的部分内容及不当之处进行了修改。

本书这次修订,我们保留了本教材的叙述风格和可读性特点,主要改动有以下几方面:

(1) 内容方面:将第3版每章正文中陈旧的内容用最新的资料来取代;(2) 习题与实践训练方面:将第3版中的判断题、选择题、应用能力训练题等进行调整和补充,以国际、国内实际经济生活数据和最近的统计信息资料为基础,进行了更新;(3) 统计学应用案例方面:为了强调统计学在不同行业实践中的应用,对每章课后的“本章案例”精心挑选,除第6、9章重新编写外,其余全部更新为2015年实际资料。因此,修订后的教材具有理论阐述简洁、清晰且准确,方法案例讲述新颖、鲜活又实用的特点。

本书由阮红伟担任主编,康燕燕担任副主编,李莉莉主审。各章撰写与修订的分工是:第1章、第2章、第5章、第7章和第8章由阮红伟编写,第3章和第9章由康燕燕编写,第4章和第6章由王希兴编写。全书结构体系及统编定稿由主编完成。

本书在编写过程中,参考和借鉴了大量的国内外同类著作、文献及报刊资料,值此出版之际,特向这些作者表示诚挚的感谢!同时,感谢为本次修订提出宝贵意见的教师和读者!感谢电子工业出版社的相关工作人员对本书出版的大力支持!

由于作者水平有限,书中的错误和疏漏之处在所难免,恳请读者给予批评和指正,以便进一步改进和完善。

阮红伟

Ruanhwqdu@163.com

目 录

第 1 章 总论	1
1.1 统计学的研究对象	1
1.1.1 统计的含义	1
1.1.2 统计研究对象的特点	2
1.1.3 统计的分类	3
1.2 统计工作过程与研究方法	4
1.2.1 统计工作过程	4
1.2.2 统计研究方法	6
1.3 统计学的基本概念	7
1.3.1 统计总体与样本	7
1.3.2 标志与指标	8
1.4 数据的计量尺度	10
1.4.1 定类尺度	11
1.4.2 定序尺度	11
1.4.3 定距尺度	11
1.4.4 定比尺度	12
1.4.5 四种计量尺度的比较	12
统计术语	13
重点知识梳理	13
习题与实践训练	14
本章案例	18
第 2 章 统计调查	21
2.1 统计调查的意义和种类	21
2.1.1 统计调查的意义和特点	21
2.1.2 统计调查的作用和要求	22
2.1.3 统计调查的种类	23
2.2 统计调查方案	24
2.3 统计调查方式	28
2.3.1 普查	29

2.3.2 抽样调查	30
2.3.3 统计报表	31
2.3.4 重点调查	33
2.3.5 典型调查	33
2.4 统计调查的方法和技巧	34
2.4.1 统计调查方法	34
2.4.2 统计调查技巧	36
2.5 Excel 在数据搜集中的应用	38
统计术语	40
重点知识梳理	41
习题与实践训练	41
本章案例	46
第3章 统计整理	50
3.1 统计整理的意义和内容	50
3.1.1 统计整理的意义	50
3.1.2 统计整理的内容	51
3.2 统计分组	52
3.2.1 统计分组的概念和作用	52
3.2.2 统计分组的种类	54
3.2.3 分组标志选择及界限的确定	56
3.2.4 统计分组的方法	57
3.3 分配数列	58
3.3.1 分配数列的意义和种类	58
3.3.2 变量数列的编制	59
3.4 统计图表	66
3.4.1 统计表	66
3.4.2 统计图	70
3.5 Excel 在数据整理中的应用	73
3.5.1 利用 Excel 进行统计分组	73
3.5.2 利用 Excel 绘制统计图	77
统计术语	79
重点知识梳理	80
习题与实践训练	80
本章案例	85
第4章 总量指标和相对指标	90
4.1 总量指标	90

4.1.1 总量指标的意义与种类	90
4.1.2 总量指标的计量单位	91
4.1.3 总量指标的计算和应用	93
4.2 相对指标	94
4.2.1 相对指标的意义与种类	94
4.2.2 相对指标的计算	95
4.2.3 相对指标的应用	103
4.3 Excel 在总量指标和相对指标中的应用	104
4.3.1 Excel 在总量指标中的应用	104
4.3.2 Excel 在相对指标中的应用	105
统计术语	105
重点知识梳理	106
习题与实践训练	106
本章案例	112
第 5 章 平均指标和标志变异指标	115
5.1 平均指标的意义和种类	115
5.1.1 平均指标的意义和作用	115
5.1.2 平均指标的种类	117
5.2 数值平均数	117
5.2.1 算术平均数	117
5.2.2 调和平均数	120
5.2.3 几何平均数	122
5.3 位置平均数	123
5.3.1 众数	123
5.3.2 中位数和四分位数	125
5.3.3 应用平均指标要注意的问题	129
5.4 标志变异指标	131
5.4.1 标志变异指标的意义和作用	131
5.4.2 标志变异指标的计算及应用	132
5.5 Excel 在平均指标和标志变异指标中的应用	138
5.5.1 Excel 在平均指标中的应用	138
5.5.2 Excel 在标志变异指标中的应用	140
5.5.3 Excel 描述统计工具应用	143
统计术语	145
重点知识梳理	146
习题与实践训练	146
本章案例	153

第6章 抽样推断	154
6.1 抽样推断的基本概念	155
6.1.1 总体和样本	155
6.1.2 参数和统计量	156
6.1.3 样本容量和样本个数	159
6.1.4 重复抽样和不重复抽样	159
6.2 抽样误差	160
6.2.1 抽样误差的概念	160
6.2.2 抽样平均误差	161
6.2.3 抽样极限误差	164
6.2.4 抽样极限误差的概率度	164
6.3 抽样推断的方法	165
6.3.1 抽样估计	165
6.3.2 样本容量的确定	168
6.3.3 抽样的组织形式	169
6.4 参数假设检验	177
6.4.1 假设检验的基本概念	177
6.4.2 假设检验的步骤	177
6.4.3 假设检验中的两类错误	179
6.4.4 总体均值和总体成数检验	179
6.5 Excel 在抽样推断中的应用	182
6.5.1 利用 Excel 进行区间估计	182
6.5.2 利用 Excel 进行假设检验	183
统计术语	184
重点知识梳理	185
习题与实践训练	185
本章案例	189
第7章 时间数列	192
7.1 时间数列的概念与种类	192
7.1.1 时间数列的概念	192
7.1.2 时间数列的种类	193
7.1.3 时间数列的编制原则	194
7.2 时间数列的水平指标	196
7.2.1 发展水平	196
7.2.2 平均发展水平	196
7.2.3 增长量	203
7.2.4 平均增长量	204

7.3 时间数列的速度指标	205
7.3.1 发展速度	205
7.3.2 增长速度	206
7.3.3 平均发展速度	207
7.3.4 平均增长速度	210
7.4 时间数列趋势分析预测	211
7.4.1 长期趋势分析预测	212
7.4.2 季节变动分析预测	219
7.5 利用 Excel 进行时间数列分析	222
7.5.1 利用 Excel 进行水平分析与速度分析	222
7.5.2 利用 Excel 进行长期趋势分析	224
7.5.3 利用 Excel 进行季节变动分析	226
统计术语	230
重点知识梳理	230
习题与实践训练	230
本章案例	238
第 8 章 统计指数	240
8.1 统计指数的概念和种类	240
8.1.1 统计指数的概念	240
8.1.2 统计指数的种类	241
8.2 综合指数	243
8.2.1 数量指标综合指数	244
8.2.2 质量指标综合指数	247
8.3 平均指数	249
8.3.1 加权算术平均指数	249
8.3.2 加权调和平均指数	251
8.4 指数体系及因素分析	252
8.4.1 指数体系的含义与作用	252
8.4.2 因素分析应用举例	253
8.5 常用价格指数简介	262
8.5.1 消费者价格指数	262
8.5.2 股票价格指数	265
8.6 Excel 在统计指数分析中的应用	268
8.6.1 利用 Excel 进行指数计算	268
8.6.2 利用 Excel 进行因素分析	269
统计术语	270
重点知识梳理	271



习题与实践训练	271
本章案例	276
第9章 相关分析与回归分析	278
9.1 相关分析	278
9.1.1 相关关系的概念	278
9.1.2 相关关系的种类	279
9.1.3 相关图表	281
9.1.4 相关系数	283
9.2 回归分析	284
9.2.1 回归分析的意义	284
9.2.2 回归分析的特点	285
9.2.3 一元线性回归方程	285
9.2.4 估计标准误差	288
9.2.5 判定系数	289
9.3 应用相关分析和回归分析应注意的问题	290
9.3.1 在定性分析的基础上进行定量分析	290
9.3.2 要注意现象质的界限及相关关系作用的范围	290
9.3.3 要将各种分析指标结合应用	290
9.3.4 要尽可能使用大样本材料	291
9.4 Excel 在相关回归分析中的应用	291
9.4.1 利用 Excel 进行相关分析	291
9.4.2 利用 Excel 进行回归分析	293
统计术语	294
重点知识梳理	295
习题与实践训练	295
本章案例	301
附录 A 【习题与实践训练】答案	303
附录 B 正态分布概率表	316
附录 C 随机数表（摘录）	318
附录 D t-分布临界值表	319
参考文献	321

总论

学习要点

- 统计的含义、研究对象和特点。
- 统计的基本概念及各概念之间的区别与联系。
- 统计研究的基本方法和一般过程。

对于“统计”一词，相信大家都曾听说过。提到统计就会使人联想到数据。有人估算，随着科技的发展，数据正在以每年超过 50% 的速度增长。这意味着每二十个月左右的时间世界上的数据就会翻一番。联想到电子商务、网上社交平台、网络媒体、互联网、物联网等的发展，数据将会像土地、石油和资本一样，成为经济运行中的根本性资源。

数据静静地待在我们生活的每一个角落，如同果园里丰收的果子等待着我们去采摘。对大多数人来说，真正有意思的并不是数据本身，而是数据背后蕴含的信息。人们想知道，面对浩繁的数据该怎样选取、如何整理、如何进行统计分析，以得出有意义的结论。若能掌握这套流程的基本理论和方法，商务、经济、管理等领域中的数据将会得到更有效的开发和利用。难怪 Google 首席经济学家 Hal Varian 在 2012 年会说，统计将是未来十年内最迷人的职业。

统计是人们认识社会非常重要的一个工具，本章将从总体上阐述统计研究的一般问题。

1.1 统计学的研究对象

1.1.1 统计的含义

统计（statistics）一词在不同的场合有 3 种含义：统计工作、统计资料、统计学。

统计工作（statistical operation）是指具体从事的统计设计、资料搜集与整理、分析预测及提供各种统计资料的实践活动的总称。例如，计算反映某公司 2015 年经营成果的各种指标，并且进行分析评价的工作过程就是统计工作。

统计资料（statistical data）是指在统计工作过程中所获得的以统计数据表现的信息资料。例如，反映某公司 2015 年经营成果的各种指标就是统计资料。



统计学（statistics）是阐述正确指导统计活动科学原理和方法的学科体系。从广义上讲，统计学是包括自然科学和社会科学在内的统计科学理论的总和。本书专门阐述作为社会科学分支的统计学理论和方法，即社会经济统计学，主要论述对社会经济现象如何进行统计设计、统计调查和统计整理及分析统计资料的理论和方法。

“统计”一词的3种含义有着密切联系。统计工作与统计资料是工作过程与成果的关系，统计学与统计工作是理论与实践的关系。因此，“统计”一词是统计工作、统计资料、统计学的综合概括，是统计的过程与结果、理论与实践的辩证与统一。

统计作为搜集客观实际资料的社会实践活动已经有几千年的历史了，但“统计”用语的出现只有不到300年的时间。在17世纪早期，人们就开始收集与人口相关的数据。以此为基础，科学家（大多数是数学家，此外是物理学家和生物学家）需要发展特定的工具来回答特定的问题。18世纪，德国政治学家阿亨瓦尔（G.Achenwall，1719—1772）最早把“国势学”定名为Statistik，即统计学。此后，各国相继沿用“统计”这个词，并把它译成各国文字。1903年，我国学者钮永建、林卓南等翻译出版了日本横山雅男所著的《统计讲义录》一书，“统计”这个词才传到我国。1907年，彭祖植编写的《统计学》是我国最早的一本统计学书籍。“统计”一词在我国从此就成了记述国家和社会状况的数量表现和数量关系的总称。

1.1.2 统计研究对象的特点

社会经济统计的研究对象是大量社会经济现象总体的数量方面，其特点可概括为如下三个方面。

1. 数量性

统计的研究对象并不是社会经济现象总体的各个方面，而是研究其数量方面，即以数据为依据具体说明社会经济现象总体的数量特征、数量关系和数量界限。例如，社会人口的数量及构成，社会财富的数量及分配，经济发展的规模和速度，人民群众物质文化水平现状及变化状况等。利用反映这些现象数量方面的各种统计数据，能说明社会经济现象发展状况、发展变化关系及预测未来，生动形象，具有较强的说服力。

统计研究对象的数量性，是统计区别于其他社会经济调查研究活动的根本特点。但是必须指出，统计是在质与量的辩证统一中研究社会经济现象总体的数量方面的。社会经济统计不是“纯数量”的研究，这是统计与数学的重要区别。统计反映的数量具有具体性，而数学反映的数量具有抽象性。例如，反映社会经济现象特征的统计数据，都是明确规定了具体的时间、地点条件下的数量。2014年（时间）我国（地点）的国内生产总值为636 463亿元，这个数量就是2014年我国的国内生产总值数量，而不是其他国家、其他时间条件下的数量。如果抽掉具体的内容，不是在一定时间、地点和条件下进行研究，那就不能说明任何问题，也就不能称其为统计，其数字也就不是统计数字。

统计的数量性与会计学反映的数量也有区别。会计学主要研究现象的价值量，而统计学不但研究价值量，还要研究实物量和劳动量；会计核算主要描述数量表现，统计研究不但描述数

量表现，还要研究数量关系和数量界限。

此外，大数据是以容量大、类型多、存取速度快、应用价值高为主要特征的数据集合，其数据格式多样，图表、文本、声音、图像、指纹、网络日志、视频、图片、地理位置信息等类型繁多。因此，统计的数量性与大数据有区别。

大数据正快速发展为对数量巨大、来源分散、格式多样的数据进行采集、存储和关联分析，从中发现新知识、创造新价值、提升新能力的新一代信息技术和服务业态。全球范围内，运用大数据推动经济发展、完善社会治理、提升政府服务和监管能力正成为趋势。大数据正日益对全球生产、流通、分配、消费活动及经济运行机制、社会生活方式和国家治理能力产生重要影响，它的迅猛发展为统计学提高统计质量、降低统计成本、扩大统计作用领域及延伸统计学科体系等提供了巨大的机遇。

另外，统计与其他学科也有关系，统计可以帮助其他学科探索其学科内在的数量规律性，而对这种规律性解释的深入研究则由各学科完成。

2. 总体性

统计的研究对象不是个体现象的数量方面，而是社会经济现象总体的数量方面。统计的研究结果都是描述总体特征的，但是总体是由个体所构成的，要认识社会经济现象总体，必须从个体入手进行调查研究。“研究个体”是过程和手段，“反映总体”是结果和目的。例如，要研究某班全体同学的某门课程的平均成绩，必须搜集每位同学的成绩，然后进行汇总计算，所得平均分（如 80 分）并非指某位同学的成绩，而是代表全班总体的水平。

3. 社会性

统计的研究对象是社会经济现象，它包括人类活动的各种条件，如社会条件、自然条件；包括人类各种活动的过程和结果，如生产活动、交换活动和消费活动等。因此，社会经济统计在研究社会经济现象时，不是孤立地进行的，而是要联系有关社会经济现象进行全面具体的分析，同时也要联系有关自然现象与技术因素等方面进行研究，即具有明显的社会性。这样才能说明现象变化的原因与过程，科学认识社会经济现象。

1.1.3 统计的分类

统计学的内容十分丰富，研究和应用的领域非常广泛。从统计教育的角度，统计学的分类大致有以下两种。

1. 理论统计学与应用统计学

理论统计学（theoretical statistics）主要是指统计学中关于数据的搜集、整理和分析的最基本的原理、原则和方法，一般既适用于社会经济现象数量特征的观察和分析，也适用于自然现象和科学实验数据的分析研究，它是统计学应用于各种领域的理论基础。从这个角度讲，统计学是一门通用的方法论科学。



应用统计学（*applied statistics*）是运用于某一特定领域的统计理论与方法。统计学是一门数据科学。由于在自然科学、社会科学的所有研究和实际工作中，都要通过数据来分析和解决问题，统计方法的应用就自然而然地扩展到几乎所有的研究领域，形成了各种应用统计学。例如，统计方法在医药领域的应用就形成了医药统计学，统计方法在人口领域的应用就形成了人口统计学，统计方法在工业领域的应用就形成了工业统计学，统计方法在经济、贸易、管理、商务等领域的应用就形成了相应的经济统计学、贸易统计学、管理统计学、商务统计学等。应用统计学的不同分支所应用的基本原理和方法是一样的，但由于每个领域都有其特殊性，所以统计方法在不同领域的应用具有不同的特点。

2. 描述统计学与推断统计学

描述统计学（*descriptive statistics*）研究如何对客观现象的数量特征进行观察、搜集数据进行计量并给予概括和表述。具体说，描述统计学的主要内容包括：确定所要研究的数量特征及其计量层次；设计用来说明现象的数量特征的统计指标；搜集数据并对数据进行整理；计算统计指标并用图表表示。例如，描述某班级总人数，某次考试的平均成绩；描述“十一”黄金周期间，北京故宫旅游景点总收入；描述2000—2015年，深圳市地区生产总值平均每年增长速度等。

推断统计学（*inferential statistics*）研究如何从总体中抽取部分样本，并根据样本数据去推断总体的数量特征。例如，公司在准备投入新产品时，通常需要估计相关市场上的消费者偏好。这可以通过对随机抽取的若干家庭进行市场调查实现，调查的结果可以作为总体消费者偏好的估计。当你从供应商那里收到大批装载的货物时，你肯定希望确保货物的质量达到合同要求。如果进行全面的质量检验，将花费大量成本，甚至某些破坏性检查将会使产品报废，而通过随机抽取一定样本进行检测以推断出总体的质量，则可使问题得以解决。为了保证这种推断的精确度和可靠度，推断统计学需要研究样本的抽取方式方法、样本的大小、样本的分布、样本估计量的选择、对总体特征进行估计或推断的方法、误差的计算和控制等问题。

描述统计学和推断统计学，一方面反映了统计发展的前后两个阶段，另一方面也反映了统计方法研究和探索事物内在规律性的先后两个过程。为达到统计研究目的，如果所搜集到的是总体数据（普查），那么，采用描述统计方法可直接达到研究目的；如果所搜集到的数据只是所研究总体的一部分数据，即样本数据，则采用描述统计方法得出样本的数量特征后，还必须用推断统计方法根据样本整理出的信息对总体做出科学的推断。可见，描述统计学是整个统计学的基础，而推断统计学是现代统计学的核心。

1.2 统计工作过程与研究方法

1.2.1 统计工作过程

统计是认识社会的工具。统计对社会经济现象的研究过程，也就是对整个社会经济现象的认识过程。统计的认识活动与人类其他的认识活动一样，是一个由浅入深、由表及里、不断深

化、永无止境的过程。随着客观事物的不断发展变化,统计认识活动需要不断进行。一般而言,一个完整的统计工作过程可以分为四个主要阶段,即统计设计、统计调查、统计整理和统计分析。

1. 统计设计

统计设计是统计工作的第一个阶段。它是根据统计研究对象的性质特点和统计研究目的,事先对统计工作的各方面和各环节所进行的通盘考虑和安排。

统计设计的主要内容如下:

(1) 统计指标和指标体系的设计。这是统计设计的核心内容。统计设计要根据统计的任务、目的及研究对象的特点,选择能反映现象本质特征的指标组成指标体系;同时,还要考虑指标间的相互联系,明确指标口径范围、计算方法和重要的分组等。

(2) 搜集整理资料方法的设计。根据统计的目的和任务确定适当的搜集资料方法,如普查、抽样调查、重点调查等。统计整理同样也有多种方法需要预先选择好。

(3) 统计工作各部门和各阶段的各种保证条件的要求。统计各部门的协调、统计各阶段的联系、统计活动需要的人员和经费等,都需要预先考虑安排好。

(4) 具体实施方案的设计。在考虑了上述三个方面之后,要求具体安排各个环节,提出日程表和工作进度,以便监督实施。

统计设计的结果表现为各种设计文案,主要包括:各种标准、制度、规定、方案和方法等。例如,统计分类标准和目录、统计指标体系、统计报表制度、统计调查方案、统计整理和汇总方案等。

统计设计在统计工作中起着重要作用。统计研究对象是社会经济现象总体,往往涉及面广,工作量大,投入的人力、物力、财力较多,对经济活动影响较大,这就要求统计工作要高度集中统一。无论是统计总体范围、统计指标口径和计算方法,还是统计分类和分组标准,都必须统一。因此,只有事先进行统计设计,才能做到统一认识、统一步骤、统一行动,使统计工作有秩序地协调进行,保证统计工作的质量。

2. 统计调查

统计调查是搜集统计资料的工作过程。它的任务是根据统计设计的要求,采用科学方法,针对社会经济现象进行有计划、有组织地搜集统计资料工作。

统计调查是统计的基础工作,是统计活动定量认识的起点,非常重要。统计调查阶段所搜集的资料是否准确、及时、全面、系统,将直接关系到统计整理的好坏,影响统计分析的结论,决定统计工作的质量。

统计调查的结果表现为各种调查表、登记表等原始数据的反映。

3. 统计整理

统计整理是根据统计研究目的,将统计调查所搜集到的原始资料进行科学的分类和汇总,使之系统化、条理化,为统计分析提供能描述现象总体数量综合特征资料的工作过程。

统计整理是统计工作的中间环节。统计整理能帮助我们由对社会经济现象的个体认识过渡到对总体的认识，由感性认识过渡到理性认识。统计整理是统计调查阶段的深入和继续，又是统计分析阶段的基础和前提，起着承上启下的作用。

统计整理的结果表现为各种整理表、统计图等。

4. 统计分析

统计分析是对经过加工整理的统计资料进行分析研究，采用各种统计分析方法，计算各种统计分析指标，目的是认识和揭示现象的本质和规律，得出科学结论，进而进行预测或作为决策依据的工作过程。

统计分析是统计工作的最后阶段，属于认识的理性阶段，是统计研究的决定性环节。有了科学的统计分析结论，才能充分发挥统计的信息、咨询和监督职能。

统计工作过程的四个阶段是相互联系、不可分割、依序进行的。在实际工作中，只有做好每一阶段的工作，才能保证整个统计工作高质、高效地完成。

1.2.2 统计研究方法

在长期的统计实践活动中，人们根据统计研究对象的特点和研究目的的需要，总结和创造出了一系列统计科学研究方法，如大量观察法、统计分组法、综合指标法、动态分析法、指数分析法、抽样推断法、相关分析法、统计预测法等。这一系列方法的有机结合构成了统计特有的研究方法体系，其中，大量观察法、统计分组法、综合指标法是最基本、最主要的研究方法。准确把握这些方法的基本思想和精神实质，对于搞好统计工作具有十分重要的意义。

1. 大量观察法

大量观察法就是对现象总体中的全部或足够多数的单位进行观察并加以综合分析的方法。统计学研究的是社会总体而不是个别的社会现象，由于社会现象的复杂性和总体性，必须对总体进行大量观察和分析，研究其内在联系，方能反映社会现象的规律。例如，人口现象中的男女比例问题。就单独一个家庭来观察，新生婴儿的性别可能是男性，也可能是女性。从表面上看，新生婴儿的性别比例似乎没有什么规律可循，但如果对大量家庭的新生婴儿进行观察，就会发现新生婴儿中男孩略多于女孩，大致为每出生100个女孩相应地就有103~107个男孩出生。这个性别比例103~107:100就是新生婴儿性别比的数量规律。古今中外这一比例都大致相同，这是由人类自然发展的内在规律所决定的。尽管从新生婴儿来看，男性婴儿略多于女性，似乎并不平衡，但由于男性的死亡率高于女性，到中年时，男、女人数就大体相当了。进入中老年后，男性死亡率仍然高于女性，导致男性的平均预期寿命比女性短，老年男性反而少于老年女性。只有对整个社会的所有家庭或足够多数的家庭成员进行调查，才能准确揭示人口现象中男女比例的一般特征和规律性。大量观察法是社会经济学的基本观察方法。

2. 统计分组法

统计分组法就是根据现象的特点和统计研究的目的，将现象按不同类型或不同性质划分成

若干个部分的统计方法。社会经济现象总体是由具有某种同质性的许多单位组成的群体，但由于在不同总体范围内的单位之间具有一定差别，因此有必要进行统计分组，以区分社会经济现象的不同类型和形态。例如，在社会人口这一统计总体中，就存在着年龄、学历、职业和收入上的种种差别。统计分组把总体内不同性质的单位区别开来，使性质相同的单位归在一个组内，可以区分现象的类型，反映总体的结构，揭示现象间的依存关系，从而使得对统计总体的数量表现、数量关系和数量界限的研究更加深入和透彻。

3. 综合指标法

综合指标法是指在统计研究中运用总量指标、相对指标、平均指标等综合指标对大量社会经济现象的数量方面进行综合分析，概括地表明社会经济现象总体数量特征的方法。统计作为认识社会的工具，必须对社会现象进行全面、科学的描述。总量指标是数量指标，侧重反映事物“量”的特征；相对指标和平均指标是质量指标，侧重反映事物“质”的特征。在统计研究中，只有综合运用这些指标，才能做到从“质”与“量”两方面反映问题，多角度、全方位、科学、公正地反映现象总体的特征。

1.3 统计学的基本概念

统计学中有许多概念，其中有几个概念是我们经常要用到的基本概念。下面对这些基本概念进行介绍。

1.3.1 统计总体与样本

1. 统计总体

统计总体简称总体 (population)，是根据统计研究目的确定的所研究对象的全体。它是由客观存在的、具有某种共同性质的许多个体组成的。例如，全国总人口构成的整体、上半年某公司销售的彩电构成的总体、2016年2月份某地发生的交通事故构成的总体等。组成总体的每个个体称为总体单位。在由全国总人口构成的总体中，每一个人就是总体单位；在上半年某公司销售的彩电构成的总体中，每台彩电就是总体单位；在2016年2月份某地发生的交通事故构成的总体中，每一起交通事故就是总体单位。

总体具有以下三方面的特征。

(1) 同质性。即构成总体的每个个体必须具有某一方面的共性。

(2) 大量性。即总体是由许多个体所组成的，而不能只是个别或少数单位。这是因为统计研究的目的是要揭示现象的规律性，而社会现象的规律性只有在大量现象的综合汇总中才能显示出来。个别单位的现象有很大的偶然性，而总体相对稳定，表现出共同性的倾向。

(3) 差异性。即总体的各个个体之间，除必须在某一方面有共性外，在其他方面还应存在差异，这些差异是统计研究的基础。

统计总体的范围可大可小。例如，要进行全国范围的人口研究，全国总人口是统计总体，总体范围较大。而要研究某班级学生身体状况，则全班同学是统计总体，总体范围较小。由于统计研究的目的和任务不同，统计总体既可以由人（如一名学生）、物（如一台彩电）、事（如一起交通事故）构成，也可以由企事业单位（如一所学校）构成。统计总体可以分为无限总体和有限总体两类。无限总体是指包括的个体数很多，以致无法计量的总体。例如，宇宙中星球的个数、海洋鱼类等都是无限总体。社会经济统计中，经常把企业自动化生产线上大量的、连续不断生产出来的小产品或零件看成是无限总体。有限总体指由有限个个体构成的总体，如全国高等院校组成的总体、某城市全体居民组成的总体。社会经济统计中，大多数都是有限总体，既可以采用全面调查，也可采用非全面调查。而对于无限总体，只能采用非全面调查方法来推断总体情况。

统计总体与总体单位的概念是相对而言的，随着研究目的和总体范围不同而相互变化。同一研究对象，在一种情况下为总体，但在另一种情况下又变成了总体单位。例如，要研究广州市所属各辖区人口数时，广州市为统计总体，每个辖区为总体单位；而当要研究全国各城市人口数时，则全国所有城市是统计总体，广州市又成了总体单位。

2. 样本

从总体中抽取一部分个体构成的集合，称为样本（sample）。构成样本的个体的数目，称为样本量。例如，从一批袋装奶粉中随机抽取 100 袋，这 100 袋奶粉就构成了一个样本，100 袋是样本量。抽取样本的目的，就是要根据这 100 袋奶粉的检测指标去推断这一批袋装奶粉的质量。

样本的单位数相对于总体来说，虽然只是很少的一部分，但样本是从总体中产生并代表总体的特征，因此，样本所取自的总体又称为母体，样本又称为子样。

样本是统计学中非常重要的概念。样本既然是从总体中以某种方式抽取的，那么它们就具有与总体同质的数量特征，但由于每次抽取样本时客观条件可能不同，因此，即使对同一个总体，按照同一样本量抽取样本，得到的样本也是多种多样的。所以，样本具有随机性。

1.3.2 标志与指标

1. 标志

标志（mark）是反映总体单位的属性和特征的名称。例如，某班级学生构成一个统计总体，每一位学生是这个总体的总体单位，反映学生的各种特征的名称，如性别、籍贯、年龄、身高、学习成绩等，都称为总体单位的标志。

标志可分为品质标志和数量标志。品质标志是指反映总体单位“质”的特征的名称，只能用文字表示，不能用数值表示。例如，学生的性别、籍贯、民族等，都属于品质标志。数量标志是指反映总体单位“量”的特征的名称，一般用数值表示。例如，学生的年龄、身高、学习成绩等，都属于数量标志。

标志还可分为不变标志和可变标志。不变标志是指所有的总体单位共同具有的特征。例如，

在“高等院校女生总体”中，每一位女生是总体单位，性别是反映总体单位特征的标志，它在女生总体中不发生变化，即大家都是女性。此时，“性别”这个标志就是不变标志。可变标志是指在总体各单位之间存在差异的标志。例如，在“高等院校女生总体”中，身高、体重、年龄等标志在每位女同学之间都存在差异，则它们都是可变标志。不论是数量标志还是品质标志，都有可能是可变标志。统计学上把标志在各总体单位上所表现出来的差别叫作变异，所以可变标志也称为变异标志。不变标志的存在，保证了统计总体的同质性，是构成统计总体的必要条件和确定总体范围的标准；可变标志的存在保证了统计总体的差异性，是进行统计研究的兴趣和目的所在。

“标志”和“标志表现”是两个不同的概念。尽管总体各单位都具有共同的标志，但每个单位对该标志的具体表现却可能不同，而总体各单位在特定时间、地点条件下的具体表现正是统计最关心的问题。品质标志与数量标志表现各不相同。品质标志表现只能用文字、语言来描述。例如，性别是品质标志，男性或女性则是标志的具体表现；籍贯也是品质标志，其标志表现具体为山东省、江苏省、广东省等。数量标志表现是用数值来表示的。例如，身高是数量标志，165cm、170cm、175cm 等是数量标志表现；学习成绩是数量标志，具体分数 80 分、90 分等是数量标志表现。数量标志表现又称为标志值。

当一个标志既是数量标志又是可变标志时，我们称其为变量。在“高等院校女生总体”中，身高、体重、年龄等标志都是可变的数量标志，因而都可以称为变量。某一变量在各总体单位的具体表现称为变量值。例如，身高是变量，160cm、170cm、180cm 等都是变量值。可见，一个变量有若干个变量值。

变量(variable)根据其变量值的取值情况可分为连续型变量(continuous variable)和离散型变量(discrete variable)。连续型变量是指相邻的两个变量值之间可以有无限多个数值，变量值是连续不断的，取整数或取小数都具有经济含义。例如，人的身高、体重，某地的国内生产总值，固定资产投资额等都属于连续型变量。离散型变量是指两个相邻的变量之间没有小数，变量值只能用整数表示，当取小数时，变量就失去了经济含义。例如，各公司的员工数，家庭拥有的移动电话数、彩电数等，其取值只能用整数表示，因而均属于离散型变量。

2. 指标

指标是统计指标(statistical indicator)的简称，是指用于反映统计总体数量特征的概念或范畴。它是统计学中的一个重要概念，可以反映总体的规律性。一项完整的统计指标主要由四部分构成，即指标名称、指标数值、指标单位和指标条件。例如，“我国 2015 年上半年国内生产总值为 296 868 亿元”，其中“国内生产总值”是指标名称；“296 868”是指标数值；“亿元”是指标单位；“我国 2015 年上半年”是指标条件，即指标的空间及时间范围。在社会经济生活中，在统计设计阶段，时常把调查项目名称叫作统计指标，而在调查汇总之后又把指标数值叫作统计指标，这都不影响统计指标构成的完整性。统计指标具有对统计总体特征定性认识和定量认识的双重作用。

统计指标可分为数量指标和质量指标。数量指标是反映总体绝对数量多少的总量指标。例如，全国人口总数、高等院校数、国内生产总值、销售收入、出口贸易总额等，这些指标要么

是总体单位总量，要么是总体各单位标志总量，都反映了现象的总规模、总水平和工作总量。所以，数量指标也称为总量指标，用绝对数表示。质量指标是反映现象相对水平和工作质量的统计指标，如人均国内生产总值、职工平均工资、产品合格率、学生出勤率等。质量指标是总量指标的派生指标，反映现象之间的内在联系和对比关系，用相对数或平均数表示。

反映总体特征的指标叫作参数（parametric），反映样本特征的指标叫作统计量（statistic）。参数与统计量的划分只有在推断统计学中才重要。本章 1.1.3 节中提到的推断统计学，实际上就是根据样本的统计量来推断总体参数的过程。

指标和标志是既有联系又有区别的两个概念。两者的区别是：指标是反映总体特征的，而标志则是说明总体单位的；指标（不论数量指标还是质量指标）能用数量来反映，而标志中只有数量标志才能用数值来表示（品质标志不用数值表示）。两者的联系是：标志是指标的基础，没有标志和标志表现，就没有指标；指标和标志的确定并非一成不变，当总体和总体单位随统计研究目的发生变化时，指标和标志也必然随之发生相应的变化。

单个统计指标只反映总体某一个数量特征，说明现象某一侧面情况。然而，客观现象是错综复杂、相互联系的，要科学、全面地反映社会经济现象的数量特征，描述现象发展的全过程，需要将一系列有联系的统计指标有机地结合起来进行分析研究。由一系列相互联系的统计指标所组成的有机整体，称为统计指标体系。例如，为了反映公司生产经营状况，只设立利润这一项指标是不够的，应该考虑公司的投入，考虑投入与产出之比；所以，应该设立产量、增加值、职工人数、工资总额、利润、产值、劳动生产率、产值利税率、资金成本利润率等一系列指标构成的指标体系，才能反映公司经营全貌。社会经济统计指标体系大体上可分为两大类，即基本统计指标体系和专题统计指标体系。基本统计指标体系是反映国民经济和社会发展及其各个组成部分的基本情况的指标体系，它包括反映整个国民经济和社会发展的统计指标体系、各地区和各部门的统计指标体系，以及基层统计指标体系。专题统计指标体系是针对某一个经济或社会问题而制定的专项指标体系，如小康生活水平指标体系、政府绩效评估体系等。

1.4 数据的计量尺度

统计的研究对象是社会经济现象的数量方面，统计数据是统计工作的基础，没有统计数据，一切统计方法就变成了“无米之炊”。统计数据的表现丰富多彩，如有代表产品成本的数据、有代表货物运输重量的数据、有代表排位名次的数据、有代表大型超市地理方位的数据，等等。这些数据所代表的含义不同，肯定不能采用相同的方法进行计量，所以，在进行统计工作搜集数据之前，就涉及数据的计量尺度问题。

统计数据可以是定性型的，如考生的性别、产品的质量等级等，它们只对客观现象的属性加以分类；统计数据也可以是定量型的，如农民的年收入、公司的利润等，它们可对客观现象用较精确的数字或数值加以描述。按照统计的一般分类，统计数据可细分为四种计量尺度：定类尺度、定序尺度、定距尺度和定比尺度。其中，定类尺度和定序尺度是定性数据（qualitative data）；定距尺度和定比尺度是定量数据（quantitative data）。

1.4.1 定类尺度

定类尺度 (nominal level) 是一种精确水平最低、最粗略的计量尺度。它仅仅是一种标志,用以区分变量的不同值,但没有次序关系。例如,在统计问卷调查中,特别是人口统计中,经常调查人口的性别、职业的分类等。为了便于统计整理汇总,特别是为便于计算机的识别,通常给每类或每组设置一个代码,如数码 1 表示男性,数码 2 表示女性,或反过来编码,数码 1 表示女性,数码 2 表示男性。也可以用其他任意两个不同的数码代表类别。这些数码都仅仅是一个符号,不能代表次序排列,也不能用来进行运算。例如,在这里 $1+2 \neq 3$ 。

在社会经济现象中,像宗教、种族、地理方位、出生地、社会保险号码、电话号码、身份证号码等都可以用定类尺度计量。适用于分析定类数据的统计技术是有限的。然而,统计的一些更宽广的应用,如 χ^2 统计能够利用定类数据产生出有用的信息。

1.4.2 定序尺度

定序尺度 (ordinal level) 比定类尺度的精确性要高一些,定序数据表现为各类别或各组之间有一定的顺序,数据可以排序,而且可以比较大小。例如,学生的考试成绩可分为:优=1,良=2,中=3,及格=4,不及格=5。当然,也可以倒过来用 5~1 表示不同等级,信息一点也没有损失。各编码的序值,代表了考试成绩的高低差异。尽管这个差异不能准确描述差别的大小,但可以确定其顺序,亦即可对定序数据作不等式运算。例如,对考试成绩而言,如果已知“优>良”,“良>中”,则肯定有“优>中”。

在社会经济现象中,《财富》杂志对世界 500 强公司的排名,针对某服务项目按照“很好、较好、一般、较差、很差”的级别进行评价,观众对电视栏目“强烈推荐、推荐、不推荐”的评级等都需要用到定序尺度来计量。

定序数据提供了有关相对比较方面的信息,但没有提供差别程度,它与定类数据一样也不能用来进行计算。定类数据和定序数据又称为定性数据。

1.4.3 定距尺度

定距尺度 (interval level) 是比定序尺度高一层次的定量尺度。它不仅能将总体各单位分类,并使各类型具有实质性意义的排序,而且能以确切的数值反映各单位之间的距离的大小,所以它的精确性比定类尺度前进了一大步。它是定比尺度的基础,在统计数据中居于重要的地位。例如,温度 21°C 、 22°C 、 25°C 等能够按照温度高低进行排序,同时,能准确测定两个温度之间的差距。即 21°C 与 22°C 相差 1°C , 22°C 与 25°C 相差 3°C 。

定距尺度可以进行加法运算,也可以进行减法运算,它们都有实际意义。但是,定距尺度不能进行乘除运算,因为在等级序列中没有固定的、有确定意义的“0”位。0 仅仅是刻度上的另一个点,而并不意味着现象不存在。例如, 0°C 并不是最低的、起点温度,因此不能说 20°C 是 10°C 的 2 倍。实际上,将这两个摄氏温度换算为热力学温度后,这两个温度并不是 2 倍的关系。

在社会经济现象中，学生的平均成绩、以年计的年龄、城镇登记失业率、就业率、投资回报率等都是定距数据。

1.4.4 定比尺度

定比尺度（ratio level）是数据最高级的定量尺度。定比尺度具有定距尺度同样的特点，但定比尺度有绝对的 0，并且两个数据之比是有意义的。“绝对的 0”的意思是说，0 是固定的，而且“0”表示“没有”或“无”。例如，高度、重量、容积、产量、利润等均为定比数据。在这里，“0”表示该事物不存在，而且两数据的比率具有意义。我们可以说，重量 180 千克是 90 千克的 2 倍，换句话说，两个重量之比是 180:90。

在社会经济现象中，旅客乘车到达目的地的路程、产品生产周期、职工人数、各地区国内生产总值、公司的利润等都是定比数据。定比尺度在统计实践中应用非常广泛。

定比尺度与定距尺度的差别在于，定比尺度具有一个自然的零始点，数据之间的比率是有意义的，但两者通常都是在社会各领域中经过精密测算而得到的。定距数据和定比数据又称为定量数据。

1.4.5 四种计量尺度的比较

定类尺度、定序尺度、定距尺度和定比尺度的关系如图 1.1 所示。

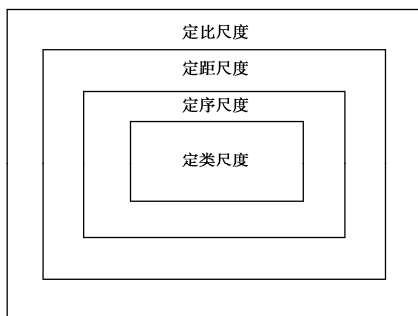


图 1.1 四种计量尺度的关系图

从最中心的定类尺度到最外围的定比尺度，它们是按照由较低的或较粗略的计量到较高的或较精密的计量排序的。一般来说，较精密的计量尺度可以转化为较粗略的计量尺度，亦即定量数据可以转化为定性数据。例如，定距尺度和定比尺度是可变成定类尺度和定序尺度的。但较粗略的尺度是不能转变为较精密的计量尺度的。而且，不同级别的计量尺度，应用范围不同。一般来说，数据的等级越高，应用范围越广泛；等级越低，应用范围越受限制。

统计尺度分四个不同的层次，只是针对客观事物量化程度和运算功能来说的，而不是指统计研究本身的高低之分。统计技术可分为参数统计和非参数统计两种。参数统计要求数据是定距数据或定比数据，如果是定类或定序数据，则应采用非参数统计。非参数统计也可以用来分析定距数据或定比数据。本书内容主要是参数统计，因此，定量数据是我们研究的重点。



统计术语

统计 statistics

统计资料 statistical data

理论统计学 theoretical statistics

描述统计学 descriptive statistics

总体 population

变量 variable

离散型变量 discrete variable

统计指标 statistical indicator

统计量 statistic

定量数据 quantitative data

定序尺度 ordinal level

定比尺度 ratio level

统计工作 statistical operation

统计学 statistics

应用统计学 applied statistics

推断统计学 inferential statistics

样本 sample

连续性变量 continuous variable

标志 mark

参数 parametric

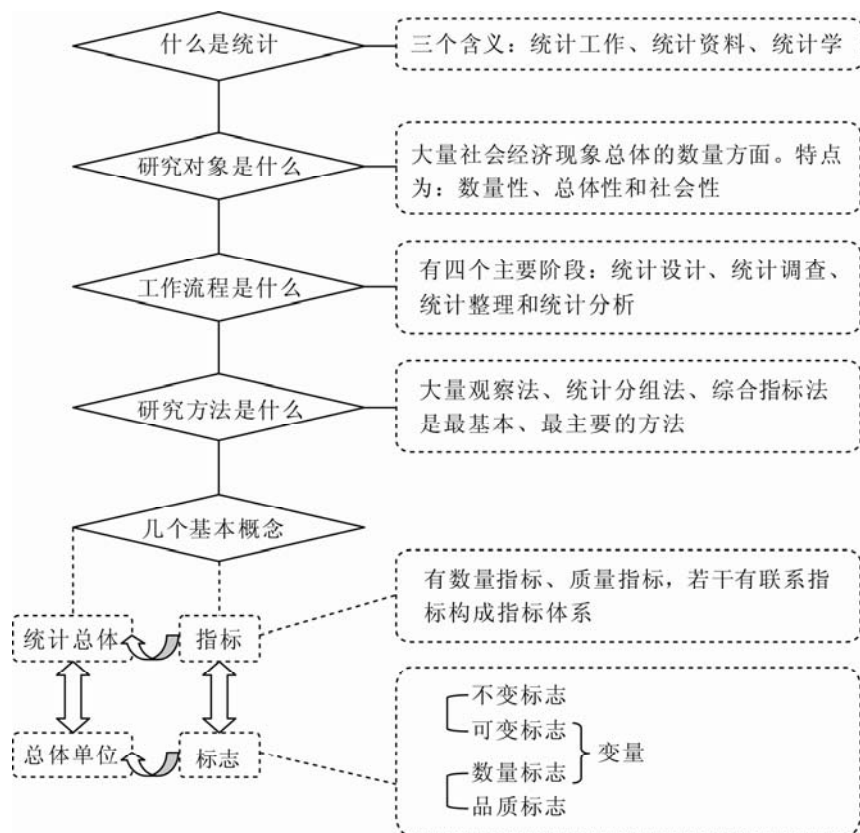
定性数据 qualitative data

定类尺度 nominal level

定距尺度 interval level



重点知识梳理



习题与实践训练

一、判断题

1. 社会经济统计的研究对象是社会经济现象总体的各个方面。 ()
2. 办公室有5位职员，年龄分别是24岁、29岁、28岁、35岁、44岁，这些年龄是5个数量标志或5个变量。 ()
3. 女性是品质标志。 ()
4. 人口的平均寿命是数量标志。 ()
5. 人口普查中，全国总人口数是统计总体。 ()
6. 数量标志可以用数值表示，质量指标不能用数值表示。 ()
7. 数量指标的表现形式是绝对数，而质量指标的表现形式是相对数和平均数。 ()
8. 只有对数量标志的标志表现进行汇总才能形成统计指标。 ()
9. 推断统计学是整个统计学的基础，而描述统计学是在此基础上发展而来的。 ()
10. 某城市每个家庭拥有的汽车数是一个离散型变量。 ()

二、单项选择题

1. 要了解100名学生的学习情况，则总体单位是 ()。
 - A. 100名学生
 - B. 每一名学生
 - C. 100名学生的学习成绩
 - D. 每名学生的学习成绩
2. 5位同学《统计学》测验成绩分别为79分、82分、67分、87分、91分，则成绩是 ()。
 - A. 品质标志
 - B. 数量指标
 - C. 变量值
 - D. 数量标志
3. 对中国女排队员的身高采用哪种计量尺度进行度量是最恰当的 ()。
 - A. 定类尺度
 - B. 定序尺度
 - C. 定距尺度
 - D. 定比尺度
4. 某公司青年员工的平均受教育年限为16.76年，这是 ()。
 - A. 数量标志
 - B. 数量指标
 - C. 品质标志
 - D. 质量指标
5. 初步核算，2015年上半年中国国内生产总值（GDP）296 868亿元，按可比价格计算，同比增长7.0%。则 ()。
 - A. 国内生产总值（GDP）296 868亿元是数量指标，增长速度7.0%是质量指标
 - B. 国内生产总值（GDP）296 868亿元是质量指标，增长速度7.0%是数量指标
 - C. 两者都是数量指标
 - D. 两者都是质量指标
6. 将公司800名员工的工资额加起来除以800，这是 ()。
 - A. 对800个标志求平均数
 - B. 对800个变量求平均数
 - C. 对800个变量值求平均数
 - D. 对800个指标求平均数
7. 某旅行社要统计“春节黄金周”该旅行社发送的游客人数和获得的净利润，则游客人数和净利润两变量是 ()。
 - A. 二者均为离散变量
 - B. 二者均为连续变量

- C. 前者为连续变量, 后者为离散变量
D. 前者为离散变量, 后者为连续变量
8. 对受教育程度采用哪种计量尺度进行度量是最恰当的 ()。
- A. 定类尺度 B. 定序尺度 C. 定距尺度 D. 定比尺度
9. 统计总体的具体特征表现为 ()。
- A. 同质性、大量性、差异性 B. 同质性、客观性、大量性
C. 数量性、大量性、差异性 D. 大量性、具体性、差异性
10. 下面的指标数值属于统计量的是 ()。
- A. 考察上一年出品的每一部商业电影的电脑记录后发现, 平均放映时间是 152 分钟
B. 调查某班全部学生, 得到平均成绩 81 分
C. 从出口的袋装茶叶中随机抽查 100 袋, 测得平均重量为 454.6 克
D. 调查某医院全体医生, 测得春节期间平均每人休息时间为 3.7 天

三、多项选择题

1. 在全国人口普查中 ()。
- A. 全国总人口数是统计总体 B. 年龄是数量标志
C. 性别是品质标志 D. 男性人口数是指标
E. 某人职业是教师是品质标志表现
2. 下列指标中是质量指标的有 ()。
- A. 人均 GDP B. 人口平均寿命 C. 物价指数
D. 城镇登记失业率 E. 第三产业增加值
3. 下列变量中是连续变量的有 ()。
- A. 公司利润 B. 网络用户数 C. 存款余额
D. 机器设备价值额 E. 学生人数
4. 要了解 100 个工业企业生产情况, 则统计指标是 ()。
- A. 100 个工业企业的工业增加值 B. 每一个工人的月工资
C. 某一个工业企业的工资总额 D. 全部工业企业的劳动生产率
E. 100 个工业企业平均增加值
5. 总体与总体单位之间有下列关系 ()。
- A. 总体与总体单位的概念是可以互换的
B. 总体可以转化为总体单位
C. 总体单位可以转化为总体
D. 总体与总体单位是固定不变的
E. 只能是总体转化为总体单位
6. 某高校在校学生 12 050 人, 从中随机抽取 120 人进行调查, 发现戴眼镜者有 71 人。在这里 ()。
- A. 在校学生 12 050 人是参数 B. 随机抽取的学生数 120 人是统计量
C. 戴眼镜人数 71 人是统计量 D. 三个数值都是统计指标
E. 三个数值都是数量指标

7. 为了解某乡镇农民对春节联欢晚会的喜爱情况, 随机抽查 500 人, 表示非常喜欢的有 298 人, 则 ()。

- A. 这是推断统计
- B. 这是描述统计
- C. 500 人是统计总体, 298 人是样本
- D. 抽查人数 500 人是统计量
- E. 喜欢春节联欢晚会人数 298 人是指标

8. 年终对员工进行质量考核, 按照优秀、良好、合格、基本合格、不合格进行评价, 则若对数据进行计量应 ()。

- A. 采用定类尺度计量
- B. 采用定序尺度计量
- C. 采用定性数据
- D. 采用定量数据
- E. 不能采用定距或定比尺度计量

9. 下面研究问题中所确定的总体单位有 ()。

- A. 研究某超市销售额, 总体单位是每一次销售行为
- B. 研究全国高等院校规模时, 总体单位是每一所高等院校
- C. 研究本次会议代表所提交的合理化建议, 总体单位是每一条建议
- D. 研究某地区粮食收获率时, 总体单位是每一亩播种面积
- E. 研究企业流水线生产的产品质量, 总体单位是每一件产品

10. 2016 年 1 月份, 对全市小学五年级学生进行身体状况研究, 则 ()。

- A. 这是推断统计
- B. 这是描述统计
- C. 全市小学五年级全部学生是统计总体
- D. 每位学生各自身高、体重数值是变量值
- E. 男生平均身高是数量标志

四、填空题

1. 统计工作与统计资料是_____关系, 统计学与统计工作是_____关系。

2. 一个完整的统计工作过程可以分为四个主要阶段, 即_____, _____, _____和_____。

3. 在统计的研究方法中, _____、_____和_____是最基本的方法。

4. 统计上把_____叫作变量。

5. 标志是反映_____的特征的, 可分为_____和_____; 而指标是反映_____的特征的, 可分为_____和_____。

6. “2014 年年末我国共有普通高等学校 2 529 所” 是一项完整的统计指标, 它主要由四部分构成, 其中“2014 年年末我国”是_____、“普通高等学校”是_____、“2 529”是_____、“所”是_____。

7. 所谓统计指标体系是指_____, 社会经济指标体系大体上可以分为两大类, 它们是_____指标体系和_____指标体系。

8. 定类数据和定序数据又称为_____，定距数据和定比数据又称为_____。
9. 样本所取自的总体又称为_____，样本又称为_____。
10. 变量根据其变量值的取值情况可分为_____和_____。

五、应用能力训练题

1. 列出下表中各总体的总体单位、数量标志（2 个）、品质标志（2 个）。

总 体	总体单位	数量标志	品质标志
大学生			
公司全体员工			
2 月份生产的汽车			
暑期销售的计算机			
上年发生的交通事故			
商业网点			
手机			
旅游景点			
“两会”期间代表们所提议案			
图书馆的藏书			

2. 当研究下述问题时，请确定四种计量尺度中最恰当的计量尺度。

研究的问题	计量尺度
上班采用的交通方式	
每周的学习时间	
学生乘车到学校的行驶路程	
昨晚 22:00 看的电视节目	
城市按照综合竞争力排名	
社会保障编号	
大学课本的价格	
校园气温	
对学校食堂的评价	
学院足球队队员的身高	
大学毕业生毕业论文质量等级	
城市进出口总额	

3. 下面是社会经济生活中常用的统计指标：

轿车生产总量、旅游收入、经济发展速度、人口出生率、安置再就业人数、城镇居民人均可支配收入、恩格尔系数

这些指标中，哪些是数量指标，哪些是质量指标？如何区分数量指标与质量指标？

4. 国务院办公厅在 2014 年 7 月印发《关于开展 2015 年全国 1%人口抽样调查的通知》（以下简称《通知》），根据《全国人口普查条例》的规定，决定于 2015 年开展全国 1%人口抽样调查。



《通知》指出，开展2015年全国1%人口抽样调查，有利于查清自2010年以来我国人口在数量、素质、结构、分布及居住等方面的变化情况，为科学制定国民经济和社会发展规划，提供科学、准确的统计信息支持。

《通知》明确，2015年全国1%人口抽样调查将在我国境内抽取约6万个调查小区，覆盖人口约1400万人。主要调查人口和住户的基本情况，内容包括：姓名、性别、年龄、民族、受教育程度、行业、职业、迁移流动、社会保障、婚姻、生育、死亡、住房情况等。调查时点为2015年11月1日零时。

（来源：人民网 2014-07-08）

分析说明上述案例中什么是总体？什么是样本？什么是变量？

5. 若对校园中大学生关心的话题进行调查研究（如大学生上网状况调查、就业分配去向及初职预期收入调查等），你能说出统计工作中所涉及的统计总体、样本、标志、指标、指标体系等概念吗？

6. 查阅5篇你专业领域的期刊文献或硕博论文，仔细阅读，重点关注其中的“数据”部分，特别是用于调查、整理和分析数据的统计过程部分。可能目前你对具体的统计过程了解还不多，请尝试着还原统计总体、总体单位、标志、变量、指标、指标体系、统计量、参数等概念，理解统计设计、统计调查、统计整理、统计分析的结果与最初的研究课题的关系。

7. 访问一个日常工作中应用统计学或者熟悉了解统计学的人，他们可能是你的亲戚、邻居、朋友、同学、学长等，询问他或她对统计学的认识，喜欢什么，不喜欢什么，统计学有用吗？在什么场合运用统计工具？怎样应用？请教他或她作为一个过来人给你本学期学习统计学的建议。

8. 登录中华人民共和国国家统计局网站（<http://www.stats.gov.cn>），你会发现“统计数据”链接，进入后尽力查找与你专业相关的数据资料；登录美国人口统计局网站（<http://www.census.gov/>）、澳大利亚统计局网站（<http://www.abs.gov.au/>）等，浏览你感兴趣的数据资料。相关的国际组织及外地官方统计机构还有：亚洲及太平洋经济社会委员会—统计司（Economic and Social Commission for Asia and the Pacific (ESCAP)—Statistics Division）、欧洲经济委员会—统计司（Economic Commission for Europe (ECE)—Statistical Division）、欧洲联盟统计局（European Union: Eurostat）、国际劳工组织（International Labour Organization (ILO)）、国际统计学会（International Statistical Institute (ISI)）、经济合作与发展组织（Organization for Economic Co-operation and Development (OECD)）、联合国总部—统计司（United Nations Headquarters—Statistics Division (UNSD)）、发布数据准则告示版（Dissemination Standards Bulletin Board (DSBB)）等。



本章案例

中国境外旅游调查报告发布出境游人数破1亿人次

第四期年度《中国游客境外旅游调查报告》揭示，“千禧一代”与高消费游客助推中国出境游市场在2014年增长20%，并首次超过1亿人次，达1.07亿人次。中国旅客表示，澳大利亚、日本及法国成为未来12个月中国游客愿望清单上的三大首选目的地，法国被认为是最受中国游

客欢迎的欧洲目的地。在中国之外的地标式景点中，中国游客有生之年最想去的景点前三位是金字塔、富士山和埃菲尔铁塔。在 2015 年的调查中，80%的中国境外游游客通过手机、台式计算机及笔记本电脑预订和规划行程，该比例在 2014 年仅为 53%。

根据最新发布的第四期《中国游客境外旅游调查报告》(Chinese International Travel Monitor，简称 CITM)，中国出境游热潮正不断席卷全球，中国出境游人数更是在 2014 年突破 1 亿人次，达 1.07 亿人次，增幅为 20%，且这一趋势仍在延续。与此同时，中国出境游旅客的整体趋势开始呈现出年轻化、思想独立、精通科技和资金充裕等特点。

以下是中国游客在未来一年最想造访的目的地，其中澳大利亚、日本和法国成为中国游客愿望清单上的三大首选目的地（见表 1）。

表 1 中国游客境外目的地愿望清单前十名

排 名	国家及地区
1	澳大利亚
2	日本
3	法国
4	中国香港
5	韩国
6	美国
7	马尔代夫
8	德国
9	泰国
10	中国台湾

通过逾 3 000 名中国境外游游客和全球 1 500 名酒店经营者的调查问卷，Hotels.com 好订网的年度《中国游客境外旅游调查报告》对中国出境游市场持续增长这一现象及对全球旅游业的影响进行了深入全面的分析。

有关中国游客的数字着实令人惊叹：中国境外游游客数量在未来四年内有望达到 1.74 亿人次，年消费总额达 2 640 亿美元，几乎等同于新加坡等发达国家的年度国民生产总值。除国内的地标以外，中国游客有生之年最想前往的景点排行中，吉萨金字塔、日本富士山和巴黎埃菲尔铁塔荣登榜单前三甲（见表 2）。

表 2 最受中国游客青睐的五大境外地标

排 名	地 标
1	吉萨金字塔（埃及）
2	富士山（日本）
3	埃菲尔铁塔（法国巴黎）
4	凡尔赛宫（法国巴黎）
5	威尼斯（意大利）

韩国成为中国游客眼中最热情好客的境外国家，紧随其后的则是泰国和日本（见表 3）。

表3 中国游客眼中最热情好客的境外目的地排名

排 名	目 的 地
1	韩国
2	泰国
3	日本
4	新加坡
5	马尔代夫
6	中国香港
7	中国台湾
8	法国
9	澳大利亚
10	中国澳门

（来源：中国旅游新闻网 2015-07-17）

统计调查

学习要点

- 统计调查的意义与种类。
- 统计调查基本方式的运用条件。
- 简单统计调查方案的设计。
- 统计调查方法和统计调查技巧。

2015 年 8 月 22 日，北京市社会心理工作联合会、北京社会心理研究所及社会科学文献出版社共同发布了《北京社会心态蓝皮书：北京社会心态分析报告（2014—2015）》。蓝皮书指出，有高达 81.9% 的居民的理想就职单位与现实工作单位不相匹配，北京居民最看重工作中的情感交流，也就是所谓的人际关系。在所有人际关系中，北京居民认为能有“公平、善解人意的上司”最重要，也就是在工作中的人际关系中最看重与上司之间的情感交流。其后依次看重的因素是尊重、与同事的关系、认可和互动机会。数据调查采取多阶段分层随机抽样法，从全北京市 16 个区县抽取了 100 个社区，严格按照 2005 年全国 1% 人口抽样所得北京人口分布特点，共发放问卷 3 200 份，回收有效问卷 2 996 份，有效率为 93.6%。（中国网 2015-08-23）

这个调查结果引发了一些人的争论，他们关心统计数据的搜集方式、统计调查方法的选用。这表明，统计越来越走近生活，统计调查知识越来越被广泛地使用。2001 年，中国加入了世界贸易组织，中国政府统计加入了国际货币基金组织（IMF）的数据公布通用系统（GDDS），这意味着我们将努力依照国际标准和通行的做法开展统计工作。我们不仅要使社会公众能够获得更多的统计信息，而且还要使大家对资料产生的过程、统计调查的方法有所了解。

2.1 统计调查的意义和种类

2.1.1 统计调查的意义和特点

统计调查（statistical survey）是根据统计研究的目的和任务，采用科学的调查方法，有计划、有组织地向客观实际搜集各种原始资料或次级资料的工作过程。

原始资料是指向总体单位直接搜集的，来源于直接的调查和科学实验的第一手统计数据（primary data）。在社会科学的研究和经济管理中，我们用调查的方法搜集必要的统计数据。例

如，为进行宏观管理决策和科学研究，必须掌握最新的人口、国内生产总值、第二产业和第三产业的基本情况以及人民生活变化情况，这就需要组织专门调查以获得国民经济管理的基本数据。国家统计局系统和国务院各部、委、局的统计系统就承担着这些调查任务。在市场经济体制下，面对市场经济激烈竞争的现实，各企业迫切需要利用统计调查方法，快速准确地掌握市场需求和居民消费等市场信息，越来越多的调查公司和网络中心等承担起专门组织调查的任务，搜集特殊的数据，满足管理和研究的需求。在自然科学和工程实际的各个研究领域，我们通常通过科学实验的方法获得统计研究的数据。例如，在医学领域，可通过临床试验数据分析某种药物或治疗方案的疗效；在农业科研中，可通过实验的方法选取最优品种和最佳的种植方式等。

次级资料是指已经经过某个部门或地区加工整理过的，即来源于别人调查和科学实验的第二手统计数据（secondary data）。对于社会上大多数的研究工作者和实际工作者来说，亲自去做直接的调查往往不可能或没必要，这时可以通过各种渠道获取别人调查或科学实验的间接数据。例如，从统计年鉴、各种报表及报纸杂志上搜集资料，从因特网上搜集到世界各国有关资料等。随着网络技术的普及，次级资料的获得将变得越来越方便。由于次级资料一般都是从原始资料过渡而来，所以本节中，我们重点介绍搜集原始资料的统计调查基本方式、方法和实施调查的方案。

统计调查是原始资料的主要来源方式，它与其他社会调查相比较，具有以下特点。

- （1）初始性。统计调查直接向调查单位搜集第一手的原始资料，因此具有初始性特点。
- （2）个体性。统计调查直接面对组成调查对象的个体进行调查，并运用个体的资料经过汇总整理、综合分析进而反映总体特性，因而具有个体性特点。
- （3）数量性。统计调查搜集的资料主要是各种数字资料，所以具有数量性特点。
- （4）大量性。统计调查要对调查对象中的足够多的个体，甚至全部个体进行调查，从而具有大量性的特点。
- （5）总体性。统计调查的目的是反映调查总体的数量特征，因此具有总体性特点。

2.1.2 统计调查的作用和要求

统计调查介于统计设计与统计整理工作过程之间，它是统计整理和统计分析的基础环节和前提条件，在整个统计研究中占有十分重要的地位。

1. 统计调查的作用

统计调查的主要作用表现在以下几个方面。

- （1）统计调查是认识社会的基本方法。
- （2）统计调查为国家管理和现代企业管理提供基本统计资料。
- （3）统计调查的资料是进行经济预测的重要依据。

(4) 统计调查是实施经济评价的重要基础。

2. 统计调查的要求

统计调查必须保证质量, 才能正确反映客观事物, 作为预测未来、实施经济评价和经济管理的依据。因此, 对统计调查的基本要求是准确性、及时性、完整性、系统性和经济性。

(1) 准确性。准确性是指统计调查取得的各项资料必须真实可靠, 符合客观实际, 按事物的本来面貌如实反映情况。统计数字的真实性是统计工作的生命。统计调查的准确性不仅是技术性问题, 而且是涉及坚持统计制度、统计法规的原则问题。我国统计立法的核心就是保障统计资料的准确性、客观性。

(2) 及时性。及时性是指在统计调查规定的时间内, 尽快提供规定的调查资料, 完成规定的各项调查任务。统计资料是进行管理决策不可缺少的依据, 而客观经济现象又是不断变化的, 因而统计数据具有很强的时效性, 如果统计资料不及时, 就成了“雨后送伞”, 失去了价值。统计资料的及时性也是一个全局性的问题, 任何一个调查单位不按规定的时间提供资料, 都会影响全面的统计工作。因此, 提高统计调查的及时性必须各单位有全局观念, 有团队协作精神, 遵守统计制度和规律。

(3) 完整性。完整性是指调查单位不重复、不遗漏, 所列调查项目的资料搜集齐全。只有齐全的统计资料, 才能比较正确地反映所研究的社会经济现象的全貌。

(4) 系统性。系统性是指综合资料中的各项统计数据应该配套, 要能从不同侧面、不同层次上对调查对象的整体进行研究, 能够从事物的内部结构和外部联系上进行对比分析。

(5) 经济性。经济性是指在满足一定准确度的要求下, 要以最小的调查费用取得所需的统计资料。通常对调查数据的准确度要求越高, 则调查的费用就越大。而任何一项统计调查总有一定的费用约束, 因此, 如果一味地强调资料的准确度, 无视经济性要求, 会造成不必要的人力、财力和物力的浪费。

这五项要求是相互结合、相互依存的, 在每一次统计调查实践中, 要根据具体情况, 综合考虑。一般而言, 要以“准”为基础, 力求在准中求快, 以尽可能小的成本取得完整、系统的资料。

2.1.3 统计调查的种类

社会经济现象错综复杂, 而且又处在不断变动中, 这就决定了统计调查方法的多样性。统计调查可以从不同的角度, 按不同的标志进行分类。

1. 按被研究总体的范围分类

按被研究总体的范围不同, 统计调查可分为全面调查和非全面调查。

全面调查是针对有限总体而言的, 指对构成调查对象的所有总体单位全部进行调查登记。例如, 2013年, 为了摸清我国第二、第三产业的家底而进行的第三次全国经济普查, 就属于全面调查。



非全面调查，又称部分调查，是在总体中选择一部分个体进行调查，从所了解的局部数据来了解总体情况。例如，为了了解某地区居民的消费水平情况，可以只搜集各个收入阶层的一部分居民消费方面的实际资料；为了掌握进出口商品的质量，可以取出一部分产品做质量检验。与非全面调查密切相关的是如何选择部分个体的问题。从总体中，恰当地选取部分个体进行调查，是统计学中最重要的问题之一。非全面调查有抽样调查、重点调查和典型调查等几个调查方法。

在国外，非全面调查（sampling）都称为抽样调查，分为概率抽样（probability sample）和非概率抽样（non-probability sample）。我们通常所说的抽样调查是概率抽样。

2. 按调查登记时间的连续性分类

按调查登记时间的连续性不同，统计调查可分为经常性调查和一次性调查。

经常性调查是随着被研究对象的变化，连续不断地进行调查登记。例如，企业的产品产量、公司的利润额、商品的销售量等，必须在观察期内连续登记。可见，连续调查的资料是说明现象的发展过程，体现现象在一定时期的总量。

一次性调查是指间隔一定时间的不连续调查。例如，商业网点数、高等学校数、商品库存量等现象，短时期内不发生什么变化，一般隔一段时间进行一次调查。因此，一次性调查并不意味着只能对现象调查一次，只是没有必要进行经常性的调查，只需间隔一定时间了解现象在某一定时点上的状况。

3. 按搜集资料的方法分类

按搜集资料的方法不同，统计调查可分为直接调查和间接调查。

直接调查（survey at first hand）是指直接搜集第一手统计资料的统计调查，如问卷调查、专家访谈、电话调查等；间接调查（survey at second hand）是指间接地搜集第二手统计资料的调查，如搜集报纸、杂志、电台、调查报告等现成统计资料。

以上各种分类是交叉的关系，而不是相互排斥的关系。例如，人口普查既是全面调查，又是一次性调查；进出口商品质量检验既是直接调查，又是非全面调查。实际工作中的每一项具体调查都可以从各个角度去认识。

2.2 统计调查方案

统计调查是一项涉及面广、程序步骤多、要求严格的科学工作。一项全国性的统计调查，往往需要动员成千上万的人协调工作才能完成。为了顺利完成调查任务，在调查之前需要设计一个调查方案，使统计调查工作有计划、有组织地进行。一个完整的统计调查方案应该包括以下内容。

1. 调查目的

调查目的就是通过统计调查所要达到的具体目标。只有确定调查目的，才能解决为什么要调查的问题，才能确定搜集资料的范围和方法，才能有效地组织统计调查工作。调查目的的确

定应该具体明确、突出中心,否则,可能产生以下结果:匆忙辛苦地进行了调查,取得的资料可能是不需要的;或者材料的口径范围不一致,无法进行汇总得到综合结论,而真正需要的材料反而没有搜集。这样不仅浪费人力、物力,而且也将延误工作。

2. 调查对象和调查单位

有了明确的调查目的,就可以确定调查对象和调查单位。调查对象是根据调查目的而确定的被调查的统计总体。“统计总体”这一概念在统计调查阶段称为调查对象,不能把调查对象理解为被调查的个体。调查单位,是构成调查对象的每一个总体单位。“总体单位”这一概念在统计调查阶段称为调查单位。调查单位是进行调查登记的标志承担者,不要把调查单位理解为从事调查工作的工作部门或单位。

在调查方案中,有时还要规定统计资料的填报单位。填报单位也叫报告单位(report unit),它是填写调查内容、提交调查资料的单位。填报单位与调查单位(survey unit)有可能一致,也有可能不一致。例如,工业企业进行产品质量检验时,调查单位是工业企业的单个产品,报告单位是工业企业,这时调查单位与填报单位是不一致的;而在工业普查中,调查单位是每一个工业企业,报告单位也是每一个工业企业,两者完全一致。

3. 调查项目和调查表

调查项目是指对调查单位所要调查的内容,是调查单位所承担的基本标志。例如,在人口普查中的调查项目有姓名、性别、年龄、民族、文化程度、职业等。在确定调查项目时应注意以下几点。

- (1) 调查项目的内容要少而精,所选项目满足调查目的需要并且能确实取得资料。
- (2) 调查项目的含义要明确、具体,切忌模棱两可。
- (3) 调查项目的设置要考虑彼此间相互联系及同类调查的纵向衔接。

确定调查项目后,应加以科学分类、排列,设计成各种调查表。调查表(survey questionnaire)是调查项目的表现形式,其作用在于能够条理清晰地表述调查内容,便于登记调查资料。

调查表由表头、表体和表脚三部分组成。表头在调查表的上方,标明调查表的名称、填报单位的名称等。表体是调查表的主体部分,由表格、调查项目等组成。表脚在调查表的下方,包括调查人员或填报人员签名、审核人员签名、填报日期等。

调查表的形式有单一表和一览表两种。单一表是在一张调查表上只登记一个调查单位的资料,可以容纳较多的调查内容,且便于分类和整理。一览表是在一张调查表上可以登记很多调查单位的资料,却不能容纳较多的调查内容。

4. 调查时间和调查期限

调查时间是指调查资料所属的时间。如果所要调查的是时期现象,调查时间就是资料所反映的起讫日期;如果调查的是时点现象,调查时间就是规定的统一标准时间。统计调查中,应明确规定调查时间,以满足统计调查的准确性要求。

调查期限是指进行调查工作的起止时间，包括搜集资料和报送资料整个工作所需的时间。统计调查中，应根据任务量的大小及人力、物力、财力等情况，尽可能地缩短调查期限，以满足统计调查的及时性要求。

例如，对某市 2015 年批发和零售业商品销售情况进行调查，即对商品销售额、批发总额、零售总额等指标进行调查，这个调查时间是时期，即 2015 年这一年。从 2016 年 1 月 1 日起开始调查，截至 2016 年 1 月 31 日将资料搜集、整理完毕，则调查期限为 1 个月。又如某管理局要求所属企业在 2016 年 1 月 15 日上报 2015 年年末从业人员资料，则调查时间是标准时点 2015 年 12 月 31 日，调查期限是 15 天。

5. 调查的组织和实施

为保证调查工作的顺利进行，在调查方案中还应该有一个周密的组织实施计划。其主要内容包括：调查工作的领导机构和办事机构、调查人员的组织和业务培训、调查的方式方法、调查经费的预算和开支方法、工作进度安排等。

例如，2013 年我国开展第三次全国经济普查，普查方案的主要内容如下。

一、普查的目的和基本原则

（一）普查的目的

摸清我国各类单位的基本情况，全面调查我国第二产业和第三产业的发展规模及布局，系统了解我国产业组织、产业结构的现状及各主要生产要素的构成，进一步查实服务业、战略性新兴产业、文化产业等相关产业及小微企业的发展状况，全面更新覆盖国民经济各行业的基本单位名录库、基础信息数据库和统计电子地理信息系统，为加强和改善宏观调控，加快经济结构战略性调整，科学制定中长期发展规划，提供全面系统、真实可靠的统计信息支持。

（二）普查的基本原则

1. 突出重点。以摸清各类单位基本情况，查实服务业、战略性新兴产业、文化产业和小微企业的底数为主，辅之以其他必要的内容。

2. 优化方式。科学设计普查业务流程，普查和抽样调查相结合，以提高普查效能，减轻基层负担。

3. 统一组织。在普查机构的集中领导下，统一设计方案、统一布置培训、统一实施调查、统一处理数据、统一发布数据。

4. 创新手段。充分运用现代信息技术，全面采用手持电子终端设备和电子地图，实现普查数据的采集、报送、处理等手段的自动化、电子化，提高普查的信息化水平。

二、普查范围、对象和时间

（一）普查范围和对象

第三次全国经济普查对我国境内从事第二产业和第三产业的全部法人单位、产业活动单位和个体经营户进行登记和调查。

根据《三次产业划分规定》(国统字〔2012〕108号),第二产业包括采矿业(不含开采辅助活动),制造业(不含金属制品、机械和设备修理业),电力、热力、燃气及水生产和供应业,建筑业;第三产业包括农、林、牧、渔服务业,开采辅助活动,金属制品、机械和设备修理业,批发和零售业,交通运输、仓储和邮政业,住宿和餐饮业,信息传输、软件和信息技术服务业,金融业,房地产业,租赁和商务服务业,科学研究和技术服务业,水利、环境和公共设施管理业,居民服务、修理和其他服务业,教育,卫生和社会工作,文化、体育和娱乐业,公共管理、社会保障和社会组织。

法人单位、产业活动单位和个体经营户按照《统计单位划分及具体办法》和普查规定的单位划分及具体处理规定进行界定。

为保证基本单位的不重不漏,结合第三次全国经济普查,对农业、林业、畜牧业和渔业的法人单位、产业活动单位进行普查登记。

(二) 普查时点和时期

普查标准时点为2013年12月31日,普查时期为2013年1月1日—12月31日。

普查登记和数据采集工作从2014年1月1日至3月31日。

三、普查方法

(一) 全面登记

对法人单位、产业活动单位和个体经营户由其主要经营地普查机构负责进行全面登记,但建筑业法人单位由其注册地普查机构负责普查登记。多法人联合体不能作为一个普查单位,应分别对每个法人单位进行登记。

(二) 联网直报与手持电子终端设备(PDA)采集相结合

对所有普查对象(军队、武警系统和保密单位除外)由普查员使用手持电子终端设备进行定位、底册信息核查和相关证照拍照。联网直报单位普查表由普查机构通过国家统计局联网直报平台布置给普查单位填报,非联网直报单位和个体经营户普查表由普查员使用手持电子终端设备采集数据。

军队、武警系统的普查登记及数据采集方式由中国人民解放军、中国人民武装警察部队经济普查机构确定;保密单位的普查登记及数据采集方式由各地普查机构与相关部门协商确定。

(三) 普查与抽样调查相结合

为取得个体经营户的经营数据,普查后抽取一定比例的个体经营户由国家调查队进行配套抽样调查。

四、普查内容和普查表

对联网直报单位、非联网直报单位和个体经营户分别设置普查内容和普查表。

为满足普查公报和年鉴的需要,设置若干普查综合表。

五、普查业务流程

第三次全国经济普查的业务流程主要包括：普查区划分与绘图，确定核查单位底册，手持电子终端设备内容加载，普查告知，普查登记，数据审核、检查和验收，数据汇总，事后质量抽查与数据评估，主要数据发布，普查成果的开发与应用 10 个阶段。

（一）普查区划分与绘图（2013 年 9 月—11 月）

（二）确定核查单位底册（2013 年 9 月—12 月上旬）

（三）手持电子终端设备内容加载（2013 年 11 月—12 月）

（四）普查告知（2013 年 12 月）

普查告知书由国务院经普办统一印制。

（五）普查登记（2014 年 1 月—3 月）

（六）数据审核、检查和验收（2014 年 1 月—4 月）

（七）数据汇总（2014 年 4 月—6 月）

（八）事后质量抽查与数据评估（2014 年 5 月—6 月）

（九）主要数据发布（2014 年 6 月—8 月）

（十）普查成果的开发与应用（2014 年 7 月—2015 年 7 月）

六、普查纪律和质量控制

（一）普查纪律

（二）普查质量控制

七、普查的组织实施

（一）全国统一领导

（二）部门分工协作

（三）地方分级负责

（四）各方共同参与

（摘自第三次全国经济普查方案，国家统计局，2013-11-13）

2.3 统计调查方式

在明确了统计调查目的、调查内容等之后，就涉及采用什么方式实施调查的问题。统计调查的方式很多，本节介绍几种常用的重要统计调查方式。

2.3.1 普查

1. 普查的意义和特点

普查(census),即指普遍调查,是依据调查任务而专门组织的一次性的全面调查。它可以取得比较准确的全面统计资料,为搞清一个国家的主要国情国力、制定重要政策和长期发展规划提供重要依据。普查是一种重要的调查方式,世界各国在反映本国综合实力的国情、国力调查中,都采用普查的方式来完成。普查有以下三个主要特点。

(1) 一次性。普查一般用来调查属于一定时点上现象的总量,这些时点现象的数量在短期内往往变动不大,不必做连续登记,只需间隔一段较长时间进行一次性的调查。而且,普查的规模大、指标多、任务重,耗费大量人力、物力和时间,不可能采用经常性调查,只能采用一次性调查。例如,人口普查工作,不可能年年搞,更不可能月月搞,我国2010年进行的第六次人口普查与2000年的第五次人口普查就间隔了10年。

(2) 时点性。普查的对象主要是时点现象,所以每次普查都有标准时点。例如,2013年第三次全国经济普查的标准时点为2013年12月31日。但是值得指出的是,普查也不排斥对某些时期现象在某段时期内数量表现的调查,如第三次全国经济普查还包括对零售批发业在2013年度的销售总额的调查等。

(3) 全面性。普查对象范围广,调查内容详细,所以它比其他任何方式的调查更能掌握全面、详尽的统计资料,具有重要的分析价值。例如,我国第一次全国经济普查仅不同种类的普查表就有近50种,涉及的调查指标近千个。内容包括了单位基本属性、从业人员、财务状况、生产经营情况、生产能力、原材料和能源消耗、科技活动情况等,为国家制定国民经济和社会发展规划,为各级部门提高决策和管理水平提供了重要信息。

2. 普查的组织实施

普查的组织形式有两种。一种是自上而下组织专门普查机构,配备一定数量的普查人员,对调查单位直接进行调查登记。例如,我国的人口普查都采用这种形式。另一种是自下而上由被调查单位填报调查表格逐级上报来实施普查,如我国的物资库存普查。但即使是后一种形式,也仍需组织普查的领导机构,配备专门人员对普查工作进行组织领导。

根据普查特点,在组织实施时应按系统工程要求,严把质量关,具体实施过程如下。

(1) 建立普查的组织领导机构。普查范围广,调查单位多,花费时间长,要求必须建立各层次组织机构,安排专门人员参加普查工作,并且职责明晰、分工明确。

(2) 确定统一的标准时间。这是指规定某一时刻或现象所属的某一段起止时间作为登记普查对象的标准时点或标准时期。这样才能保证所有调查单位的调查资料都是反映同一时间特征的,避免所搜集的资料因时间不同,汇总后不能反映客观事实而失去准确性。例如,我国第六次全国人口普查规定的标准时点为2010年11月1日零时,而相差一天,我国的人口总量就会有5万人的差异。当今社会的高速变动性使得在普查时确定统一标准时间、严格保证调查资料从属这一时间至关重要。



(3) 规定各阶段工作进度。例如，第六次全国人口普查工作分为四个阶段。从人口普查启动至2010年10月底前为准备阶段，2010年11月1日至2011年2月中旬为普查登记和快速汇总阶段，2011年2月下旬至2011年12月为数据处理阶段，2012年1月至2012年12月为普查资料的开发利用和总结表彰阶段。有了工作进度，才能保证普查按计划同时开始、按期完成，使调查资料得以及时汇总，发挥统计调查的时效性价值。

(4) 规定普查的项目和指标。普查的项目必须统一，其内容、解释口径、计算方法等一经确定，不得任意增减改变，以免影响汇总综合，降低资料质量。同一种普查，每次项目和指标应力求一致，以便更好地进行历次调查资料的对比分析，观察某种现象变化发展的情况。

(5) 制定严格质量控制方法。普查工作是一项庞大的系统工程，普查资料的质量控制是贯穿始终的重要工作。它包括普查前的质量控制、普查中的质量控制和普查后的质量控制，只有步步为营实施控制，才能保证普查资料的准确性，提高普查工作的质量。

3. 周期性普查

普查作为获取有关国情、国力资料的重要统计方法，在我国已经纳入了规范化、法制化的轨道，我国政府也制定了“以周期性普查为基础”的整体统计调查方法体系。已有的重大普查工作的统筹安排为：人口普查，逢0年份实施，每10年1次。在年号末尾为0的年份进行人口普查，也是联合国所建议和提倡的，是世界大多数国家进行人口普查所采用的时间。农业普查，逢6年份实施，每10年1次。经济普查每10年进行2次，分别在逢3、逢8的年份实施。

美国人口普查始于1790年，依美国宪法规定每10年进行一次。2010年4月1日人口普查，是美国历史上的第23次人口普查。

2.3.2 抽样调查

1. 抽样调查的意义和特点

抽样调查(sampling survey)是一种非全面调查，它是按照随机原则从总体中抽出适量单位组成样本予以调查，用样本资料推算总体数量特征的一种调查方法。

抽样调查是目前国际公认、世界各国普遍采用的一种科学有效的统计调查方法。它有几个特点。

(1) 它是非全面的“代表性”调查。抽样调查通过样本的调查来推断总体的特征，它既是非全面调查，又可以精确度很高地推算总体数量特征，这一特点使它不同于全面调查，也与其他非全面调查有显著区别。

(2) 抽样调查按照随机原则取样。所谓随机性原则就是总体中调查单位的确定完全由随机因素来决定，单位中选与否不受主观因素的影响，保证总体中每一单位都有同等的中选可能性。采用随机原理抽选样本调查单位，才能保证样本的代表性。

(3) 抽样误差是可以控制和计算的，这有利于科学地把握调查的准确性。

抽样调查法在现代统计调查中有准确性高、时效性强、经济性好、应用性广的显著优越性。它已成为统计调查方法的主体，本书将在后面的第 6 章加以详细讨论。

2. 我国的抽样调查实践

我国在 20 世纪 50 年代，在城市职工住户调查、农村经济调查中曾经采用过抽样调查，后中断。改革开放后，在借鉴国外抽样方法的基础上，结合我国实际情况，国家统计局在 20 世纪 80 年代组建城市经济调查总队、农村经济调查总队，在 1998 年成立企业调查总队，各地方也相应有成调队、农调队和企调队，他们为国家成功地进行了工业产品质量抽样检验、城市住户调查、市场抽样调查、人口抽样调查、农产量抽样调查等。

随着经济的迅猛发展，快速、高效地获取准确的统计资料已成为企业在激烈竞争的社会中生存的重要手段，抽样调查在社会经济各部门得到了非常广泛的使用，许多人通过参与调查或被调查感受到统计就在我们身边。

2.3.3 统计报表

1. 统计报表的意义

统计报表（statistical report forms）是按国家统一规定的调查文件，以基层统计的原始记录为依据，自下而上，逐级定期地提供统计资料的一种调查方式。它是一张调查表，报表中的指标项目就是调查项目。

统计报表曾经在我国高度集中的计划经济年代是一种最主要的统计调查方式。我国在多年的统计实践中，最大限度地发挥了统计报表的作用，成为世界上运用统计报表最成熟的国家之一。在当前的市场经济条件下，虽然由于统计报表费时、费力，中间环节多，不再作为获得统计资料的主要和唯一形式，但在某些方面统计报表仍有不可替代的作用。

2. 统计报表的资料来源

统计报表的资料来源于基层单位的原始记录。原始记录是通过一定表格形式对基层生产经营活动所作的最初记载，是反映社会经济活动的第一手材料。原始记录种类繁多、范围广泛。例如，企业产品产量记录、原材料入库记录、职工出勤和工作记录、商品销售记录等都是原始记录，如表 2.1 所示。

表 2.1 材料领用单

领用部门：
发料仓库
用途

单据编号：
业务日期
备注

物料编号	物料名称	规格型号	单 位	批 次	请领数量	实发数量

部门领导：

技术员：

领料员：

库管员：

从原始记录到统计报表，中间还要经过统计台账和企业内部报表。统计台账是基层单位根据核算和填制统计报表的需要，按时间顺序登记原始记录的一种账册。它是为积累和整理资料而设置的，是从原始记录到统计报表的中间环节。例如，企业每天会有许多的产品入库单，这些产品入库单就是企业的原始记录；每天把这些入库单汇总得到该产品的当日产量，按日期登记在一本账册中，即统计台账，如表 2.2 和表 2.3 所示。

将统计台账中每日产量累计，得到当月产品产量，据此填入统计报表中的产品产量。可见，原始记录、统计台账和统计报表之间联系密切、层次递进。

表 2.2 产品入库（缴库）台账

月份	厂家	产品代号	产品名称	生产批号	单位	数量	单价	金额	箱数	入库时间	出库情况	备注

表 2.3 工业企业主要经济指标台账

2015 年													单位：万元	
月份	存货	产成品	固定资产净值平均余额	负债总计	流动负债平均余额	应交增值税	销项税额	工业中间投入	平均资产总额	从业人员平均人数	从业人员劳动报酬	工业生产综合能耗	所有者权益	同期所有者权益
1 月														
2 月														
3 月														
4 月														
5 月														
6 月														
7 月														
8 月														
9 月														
10 月														
11 月														
12 月														

3. 统计报表的应用

在市场经济条件下，必要的统计报表是我国统计调查方法体系的补充。目前统计报表主要是全面调查和重点调查。在实践中，统计报表经常同各种调查方法结合起来，综合运用。

(1) 统计报表在普查中应用。普查是全面调查，其调查对象是组成社会最基本的单位。这

种从基本单位到整体的调查、搜集资料并逐级上报汇总的过程就离不开统计报表制度，只有通过报表制度的规范性，才能使普查工作严格按程序进行，保证普查资料的质量。

(2) 统计报表在其他调查中应用。在抽样调查和重点调查中，为规范调查资料的填写上报程序，也需要建立统计报表制度，统计报表在非全面调查中同样得到广泛应用。

随着我国统计信息技术现代化体系的建立，我国的统计报表在报送程序和报送手段上已发生了深刻的变革。传统的手工报表正被计算机的信息处理所替代，传统的邮寄电讯报表正被远程传输及信息网络所替代，传统的自下而上逐级上报的程序随着网络的建立也可转变为“直达车”形式。

2.3.4 重点调查

重点调查是在全部调查单位中选择一部分重点单位进行调查，借以了解总体的基本情况的一种非全面调查。所谓重点单位，是指这样一些单位，它们的数目在全部单位数中只占很小的比重，但它们的标志值的总和在总体的标志总量中却占很大的比重，通过对这部分单位进行调查，就能从数量上反映总体的基本情况。

例如，要了解我国出口商品总额的基本情况，以欧盟、美国、东盟、中国香港地区、日本、韩国及中国台湾省等几个大的贸易伙伴为重点进行调查即可完成。虽然它们在我国所有的贸易伙伴中是少数，但它们是大的贸易伙伴，贸易额占绝对大的比重。对这些重点贸易伙伴进行调查，比全面调查要省时省力，而且更能及时地了解全国出口贸易总额的基本情况。在经济变动较大情况下，我国为了及时了解全国城市商品房价格的变动趋势，每月对全国 6、7 百个城市中的 70 个大中城市的价格的变化进行调查，这种调查就是重点调查。

根据帕累托 80/20 法则，在企业经营管理中和社会生活中普遍存在着不平衡的现象，这为我们发挥重点调查作用、了解总体概况提供了广阔的实践领域。

重点调查在实际应用时应注意以下两点。

(1) 重点单位的选择不带有主观因素。它只着眼于所研究现象主要标志的总量，而与平常所说“重点单位”有区别。只要调查单位的主要标志总量不占绝对大比重，仅因为技术创新、管理先进或特殊原因而被列为重点管理的，都不列为重点调查单位的范畴。

(2) 重点调查不能精确推断总体总量。重点调查是一种非全面调查，只有当调查任务只要求掌握总体基本情况，而部分重点单位又能比较集中地反映所研究的项目和指标时，采用重点调查比较适宜。

所以，利用重点调查推断总体，只能了解总体基本概况，而不能精确推断总体总量。

2.3.5 典型调查

典型调查 (representative sampling) 是根据调查的目的任务，在对所研究的现象总体进行初

步分析的基础上，有意识地选取有代表性的典型单位进行的非全面调查。

例如，为了研究大中型国有企业改革成功的经验和失败的教训，从全部大中型国有企业中选出若干先进的典型和失败的典型，分别进行调查比较。在此基础上，认识大中型国有企业改革的规律和本质。

典型调查能否取得良好效果，关键在于正确选择典型单位。所谓典型单位，是指在同一事物中能最充分、最突出地体现总体共性的代表单位。选择典型单位时应注意，要在对调查对象总体进行全面分析，对可供选择单位进行反复对比的基础上选择代表性较强的单位。选取典型单位的方法可以灵活多样。例如，为了研究获得成功的经验或总结失败的教训，应该选择调查对象中先进或落后的作为典型单位；为了了解总体的一般数量表现或一般发展趋势，可选择中等的带有普遍性的那部分单位作为典型单位。若调查总体各单位发展条件较一致，彼此之间差异小，可选一个或几个典型单位进行调查，即通常所说“解剖麻雀”式典型调查；若各单位之间发展条件差异大，涉及问题较复杂，则应按一定标志，将总体划分为几个类型，从各类中按比例选取典型单位调查，即通常所说“划类选典”式典型调查。

典型调查在实际应用时应注意以下两点。

（1）典型单位的确定是人为选择的，取决于统计调查人员对调查对象全面分析基础上主观的判断选择。这与重点调查中重点单位的选取是不相同的。

（2）一般来说，典型调查的目的是发现问题、观察趋势，而不能不加限制地直接以调查结果来推断总体。典型调查经常用来验证全面调查（如普查）数据的质量，而且典型调查的深入了解也可以作为全面调查的补充说明。

以上是从不同角度对统计调查方式的分类，各种统计调查方式均有其不同的特点、局限性，因而各有其适用范围。在实际工作中，应按照建立以周期性普查为基础，以经常性抽样调查为主体，以必要的统计报表、重点调查、典型调查、综合分析等为补充的统计调查方法体系的要求，针对不同的调查对象、调查目的和任务，灵活运用多种调查方式，才能准确、及时、全面、系统地提供统计资料。

2.4 统计调查的方法和技巧

统计调查方法是指统计工作确定调查方式之后，向客观实际搜集资料的具体方法。既包括直接搜集第一手统计资料的方法，也包括间接搜集第二手统计资料的方法。

2.4.1 统计调查方法

1. 观察法

观察法又称直接观察法，是由调查人员亲自到现场对调查对象进行观察、记录和计量，以取得第一手统计资料的调查方法。例如，对超市购买者进行调查时，为获得出向和入向人数，

则调查员亲自到超市的出入口实际观察, 并进行记录。这种方法能够保证统计资料的相对准确性。实际上, 精神分析学的奠基人西格玛·弗洛伊德(Sigmund Freud)做得很好的一件事就是观察和描述他的病人的状况。他是机敏的观察者, 通过观察大量的病患, 运用他的才智和技能发展了第一个系统而综合的人格理论。不论世人对他的观点持何种态度, 都无法否认他是一位观察高手、一位优秀的科学家的事实。

毫无疑问, 观察法需要耗费大量的人力、物力和时间, 在商务领域中, 调查信息要求时效性较强时不宜采用; 而且对历史资料进行调查时, 也无法通过直接观察搜集资料。另外, 这种方法容易受到观察者和被观察者主观因素的影响而产生各种观察误差。例如, 当人们知道在被观察时会改变行为, 观察者在现场观察一定时间后会因疲倦产生观察误差, 这些都会降低调查资料的质量。

2. 访问法

访问法又称采访法、询问法, 是由调查人员通过口头、书面等方式向被调查者了解情况, 取得第一手资料的调查方法。例如, 上面提到的在对超市购买者进行流动人数调查的同时, 对消费者的购物金额进行调查, 调查人员在出、入口对顾客进行询问, 根据回答一一记录, 并填写相应的调查卡片或表格据以搜集统计资料的调查方法。

常用的访问法有以下三种。

(1) 面谈访问法。由调查者直接与被调查者接触, 通过有目的地当面交谈获得资料的一种方法。面谈访问调查简便灵活, 可随机应变提出问题且调查表回收率高, 但调查成本高, 调查范围有限, 且调查结果易受调查人员综合素质的影响。

(2) 电话调查法。这是一种通过电话向被调查者询问调查内容的方法。电话访问能迅速获得有关资料, 成本低, 特别是在那些不容易见面访问或被调查者不愿意接受访问的情况下, 采用电话访问有可能获得调查成功。但是, 由于电话访问对象只限于通话者, 调查范围受到极大限制, 且通话时间不宜过长, 图表、照片无法显示, 故一般适用于调查项目单一、问题简单明确、要求快速获得所需调查信息的情况。

(3) 邮寄访问法。将设计好的问卷通过邮寄方式送到被调查者手中, 说明填表要求和方法, 由被调查者填写后寄回, 以获取信息资料的方法。这种方法调查面广、成本低、访问对象填写自由方便, 但调查表的回收率低、回收时间长, 调查结果的完整性往往不易控制。

(4) 互联网访问法。是指以网络为载体, 将问卷主要通过网站和电子邮件来发放, 以获取信息的方法。互联网调查访问在 20 世纪 90 年代开始流行, 网民数量的快速增长为网上调查的可行性和广泛应用提供了基础。据 CNNIC(中国互联网络信息中心)于 2015 年 7 月发布的第 36 次互联网统计报告显示, 中国网民数量达到 6.68 亿人, 手机网民规模达 5.94 亿人。网民中, 农村网民规模达 1.86 亿人。

互联网调查访问的特点是: 组织简单、费用低廉; 调查结果的客观性高; 传播速度快并能传送多媒体问卷; 采集信息的质量可靠; 没有时空、地域限制且网络调查的周期大大缩短。但同时, 网上调查也存在着调查对象难以确认和控制、调查对象的代表性差、问卷回答率低等缺



点。在社会实践中，对热点话题的调查、广告效果的调查等经常采用网上调查。

3. 实验法

实验法指调查人员根据研究目的，通过实验对比，对调查对象的某些因素之间的因果关系及其发展变化过程，进行实验观察和分析，以取得调查资料的方法。例如，若要了解饮料配方的改变对销售量的影响情况，则需选定一个地区范围，将新旧两种配方的饮料投入市场进行试验对比，观察其销售量变化和消费者反映，获得数据作为是否采用新配方的依据。

实验法能够排除主观估计偏差，直接掌握大量第一手实际资料，且能揭示或确立现象之间的相关关系，在定量分析上具有重要作用。但由于实验对象和试验环境的选择难有充分的代表性，实验调查的结论带有一定的特殊性，且实验法对调查者要求较高，花费时间较长，故其应用范围有限。

实验法一般适用于对新设计、新包装、新价格、新配方、新广告等社会经济现象的实践效果资料搜集。

4. 报告法

报告法是由被调查单位按照调查机关的调查方案要求，及时向调查机关报告统计资料的调查方法。统计报表就属于这种方法。报告法一般针对机关团体和企事业单位，特别是某些信息需要保密的行业单位，而不是对个人或个体单位调查。下级必须按规定准确及时、全面地向上级提供统计资料，具有法律行政的强制性。

报告法在统计报告系统健全、原始记录和核算工作完整的前提下，可以保证提供资料的准确性。但报告法通过颁发调查方案来搜集资料，调查者与被调查者不直接接触，不能了解社会经济活动具体情况和影响原因，且方法烦琐，所需人力、物力、财力较多。

报告法主要应用于对无法进行直接观察、访问和试验的历史资料的搜集。

5. 文献法

文献法是调查人员根据调查方案的内容和要求，搜集文献资料的一种方法。文献包括报纸、书籍、数据表格等文字、数字文献，也包括影视、图画、磁带、唱片等声音、图像文献。

文献法与实地调查方法相比，调查受控制因素少，花少量人力、财力便可获得所需资料。这种方法不仅适用于社会经济调查，而且广泛运用于科学研究，但文献法对文献资料的真实性的鉴定工作难度大。

以上调查方法中，观察法、访问法、实验法主要用于搜集一手资料，报告法和文献法主要用于间接搜集二手资料。它们都是非常重要的统计调查方法，在社会实践中得到了广泛的应用。

2.4.2 统计调查技巧

实施统计调查，除强调运用科学的方式方法外，还要善于运用调查技巧，这对于提高调查能力和效果尤其重要。

1. 调查问卷设计技巧

调查问卷又叫询问表或调查表,它是指为统计调查所用的、以设问的方式表述问题的表格。调查问卷广泛运用于民意调查和现代市场调查中,是收集资料最重要的工具。

调查问卷中问题的设计有两种,即开放式问题和封闭式问题。开放式问题在提出问题时不提供任何答案,由被调查者自由填写。这种调查问卷,主要用于希望了解的内容比较详细的情况。封闭式问题在提出问题的同时,还必须将答案设计出来,主要用于了解被调查者的基本意向。这是现代问卷调查中采用的主要形式。

在问卷设计中,为保证问卷的回收率及调查质量,应注意以下几个方面。

(1) 问卷中问题的表述要客观、准确、具体,避免抽象、笼统和有多重含义,不能带诱导性和倾向性。例如,“您认为今年市场供应和物价怎样?”这种提问既太笼统又同时在一个问题中问两件事,有多重含义,使被调查者难以回答。

(2) 对于敏感性问题,不要直接提问。可采用假定法、转移法等。例如,“假定对人口生育不加限制,您认为多子女和独生子女哪种情况更有利于培养子女成材?”这个问题采用了假定法,这样会比直接提问更有利于被调查者回答。

(3) 问卷中问题答案的设计既要遵循互斥性原则,又要遵循完备性原则。互斥性原则是指同一问题的若干答案之间关系是相互排斥的,不能有重叠、交叉和包含等情况;完备性原则是指所排列出的答案应包括问题的全部表现,不能有遗漏。

(4) 问卷中所设计的一系列问题,要讲究排列顺序。一般先易后难,而且符合逻辑性。这样能给被调查者一种轻松、方便、流畅的感觉,以顺利完成调查工作。

(5) 问卷设计一定要通过小规模访谈进行修改。

2. 调查谈话的技巧

调查谈话是由被调查者回答调查者所提问题而获得调查信息的一种方法。其关键是谈话过程能否顺利完成。要想取得成功,调查者应熟练掌握和运用各种技巧。

(1) 重视调查谈话的开始。调查者要讲究仪表风度,会寻找共同话题,友好、亲切接近被调查者,谈话头几分钟对访谈的成败至关重要。

(2) 调查提问要简洁,问题要具体,态度要自然。按照由易到难的逻辑顺序进行。

(3) 通过采用多种方式提问和巧妙引导的方法,控制谈话过程。

(4) 注意利用表情、动作等非语言因素引导被调查者的谈话朝着既定目标发展。

(5) 做好谈话的收尾工作。实际调查中,要审时度势,尽量使谈话在最佳点上结束,要在笑声中分手,不要在分歧中告别。

2.5 Excel在数据搜集中的应用

统计学是一门应用性非常强的学科，而统计工作的每一个环节都离不开统计计算机软件的应用。典型的统计软件有 SAS、SPSS、Minitab、Statistica、Excel 等。其中，由美国微软公司开发的 Excel 电子表格软件，是办公室自动化中非常重要的一款软件。它不仅能够进行表格处理、图形分析和数据的自动处理和计算，而且其简单易用性更能为广大办公自动化人员所接受。很多企业都是依靠 Excel 进行数据管理的。本节仅介绍 Excel 在本章中的应用。

搜集数据的方法有很多，如重点调查、典型调查、统计报表、抽样调查等。抽样调查是实际中应用最广泛的一种专门调查方法。针对抽样调查，Excel 的数据分析工具中提供了一个专门的“抽样”工具，可以帮助使用者快速完成抽样工作。

以下是利用“抽样”工具完成抽样调查的操作方法。本节操作以【实例 2.1】为操作范例进行阐述。

【实例 2.1】 假定有 100 名消费者，从中选取 16 人对其工资收入进行调查。

解：在使用“抽样”工具之前，应先输入总体单位的编号，形成编号数据清单。总体单位编号，可以按照随机原则，也可以按有关标志或无关标志进行编排。

将 100 名消费者编号为 1~100 号，输入编号，形成总体各单位编号表，如图 2.1 所示。

1	11	21	31	41	51	61	71	81	91
2	12	22	32	42	52	62	72	82	92
3	13	23	33	43	53	63	73	83	93
4	14	24	34	44	54	64	74	84	94
5	15	25	35	45	55	65	75	85	95
6	16	26	36	46	56	66	76	86	96
7	17	27	37	47	57	67	77	87	97
8	18	28	38	48	58	68	78	88	98
9	19	29	39	49	59	69	79	89	99
10	20	30	40	50	60	70	80	90	100

图 2.1 总体各单位编号表

完成总体单位编号清单后，利用“抽样”工具进行抽样的具体操作如下。

第一步：单击“工具”菜单，选择“数据分析”选项，如图 2.2 所示。

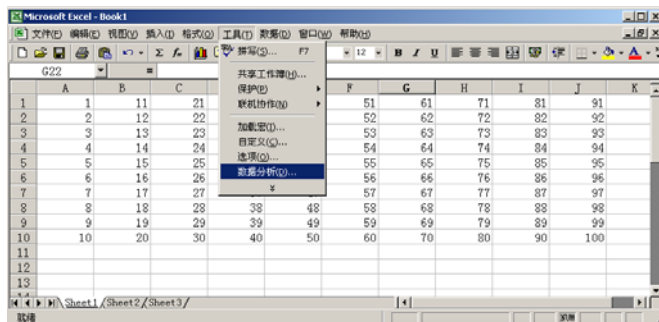


图 2.2 选择“数据分析”选项

第二步：打开“数据分析”对话框，从其对话框的“分析工具”列表中选择“抽样”，如图 2.3 所示。

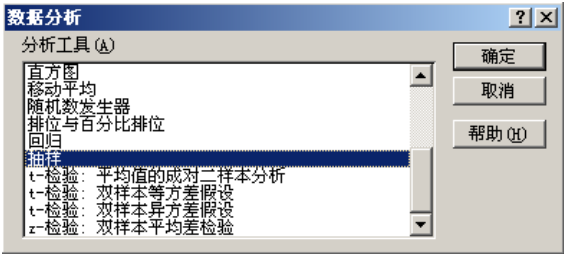


图 2.3 “数据分析”对话框

第三步：单击“确定”按钮，打开“抽样”对话框，确定输入区域、抽样方法和输出区域，如图 2.4 所示。

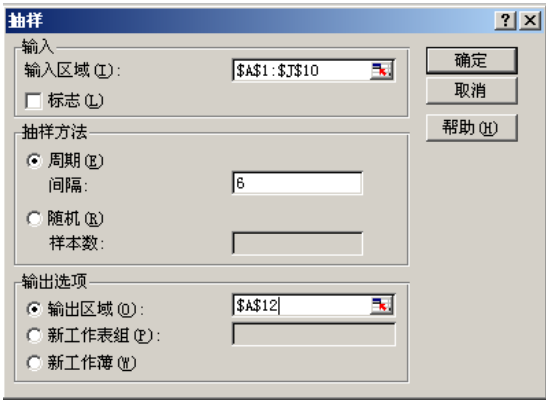


图 2.4 “抽样”对话框

（1）确定输入区域。在“抽样”对话框的“输入区域”框中输入总体单位编号所在的单元格区域。在本例中，输入区域为\$A\$1:\$J\$10，系统将从 A 列开始抽取样本，然后按顺序抽取 B 列至 J 列。如果输入区域的第一行或第一列为标志项，可单击“标志”复选框。

（2）选定抽样方法。在抽样方法项下，有周期和随机两种抽样模式。

- ① “周期”模式即所谓的等距抽样或机械抽样、系统抽样。此种抽样方法，需要确定周期间隔，周期间隔由总体单位数除以要抽取的样本单位数而求得。本例中，要在 100 个总体单位中抽取 16 个，则在“间隔”框中输入 6。
- ② “随机”模式适用于简单随机抽样、分层抽样、整群抽样和阶段抽样。简单随机抽样即纯随机抽样，只需在“样本数”框中输入要抽取的样本单位数即可。分类抽样即类型抽样，先将总体单位按某一标志分类编号，然后在每一类中随机抽取若干单位，这种抽样方法实际是分组法与随机抽样的结合。整群抽样，先将总体单位分类编号，然后按随机原则抽取若干类作为样本，对抽中的类的所有单位全部进行调查。本例所使用的编号输入方法，只适用于等距抽样和简单随机抽样。

(3) 确定输出去向。在“抽样”对话框的“输出选项”框中有三种输出去向。在“输出区域”框中输入总体单位编号所在的单元格区域。在本例中，输出区域为\$A\$12。也可以通过选择“新工作表组”或“新工作簿”将抽样结果放在新工作表或新工作簿。

第四步：单击“确定”按钮后，在指定位置给出抽样结果，如图 2.5 所示。

	A	B	C	D	E	F	G	H	I	J	K
1	1	11	21	31	41	51	61	71	81	91	
2	2	12	22	32	42	52	62	72	82	92	
3	3	13	23	33	43	53	63	73	83	93	
4	4	14	24	34	44	54	64	74	84	94	
5	5	15	25	35	45	55	65	75	85	95	
6	6	16	26	36	46	56	66	76	86	96	
7	7	17	27	37	47	57	67	77	87	97	
8	8	18	28	38	48	58	68	78	88	98	
9	9	19	29	39	49	59	69	79	89	99	
10	10	20	30	40	50	60	70	80	90	100	
11											
12											
13	6										
14	12										
15	18										
16	24										
17	30										
18	36										
19	42										
20	48										
21	54										
22	60										
23	66										
24	72										
25	78										
26	84										
27	90										
28	96										
29											

图 2.5 等距抽样结果



统计术语

统计调查 statistical survey

非全面调查 sampling

非概率抽样 non-probability sample

开放式问题 open-ended questions

统计报表 statistical report forms

电话调查 telephone surveys

调查方案 survey programmer

调查单位 survey unit

访谈 interviews

普查 census

概率抽样 probability sample

封闭式问题 closed-ended questions

抽样调查 sampling survey

网上调查 survey on the net

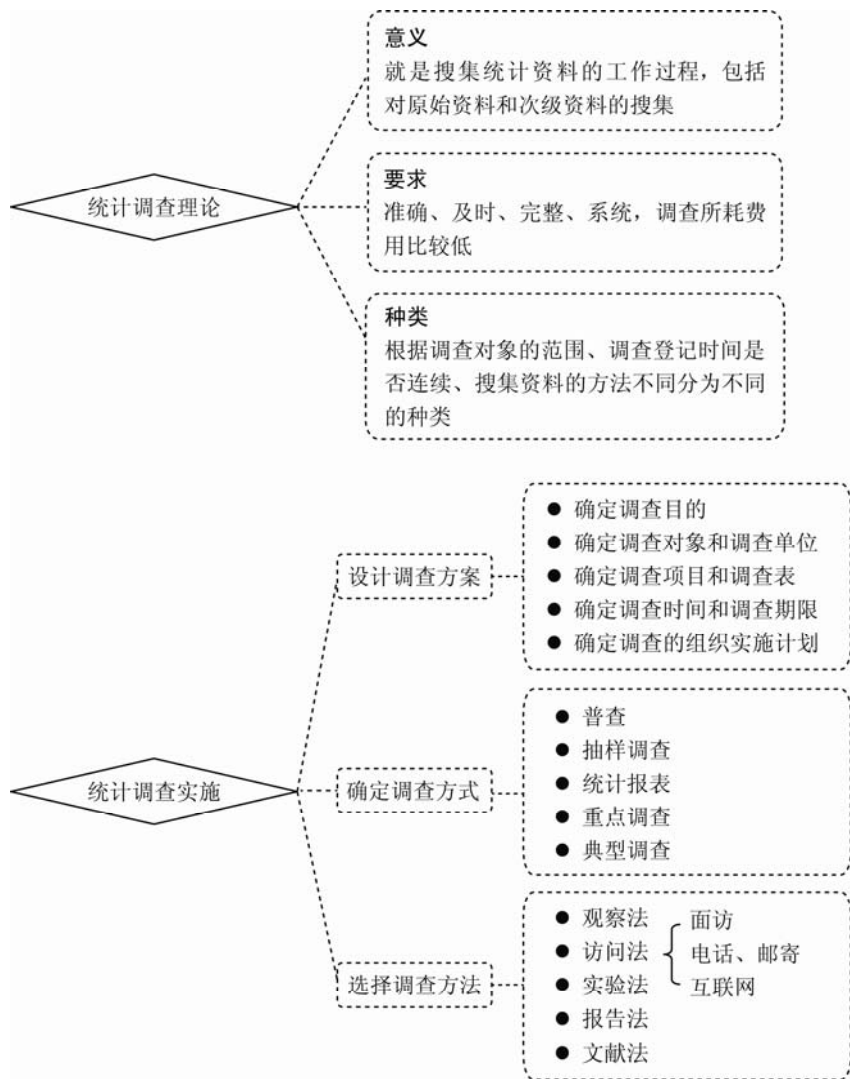
典型调查 representative sampling

调查表 survey questionnaire

报告单位 report unit



重点知识梳理



习题与实践训练

一、判断题

1. 调查单位同时又一定是填报单位。 ()
2. 我国经济普查今后每十年进行两次，因此，它是一种经常性调查方法。 ()
3. 调查时间就是调查工作所需要的时间。 ()
4. 要了解“十一”黄金周期间我国铁路旅客周转量，只需对全国几个大的铁路枢纽客运量进行调查，就可以掌握全国基本情况。这种调查属于非全面调查。 ()
5. 重点调查中的重点单位是根据进行统计调查时，当前工作的重点来确定的。 ()

6. 典型调查由于在选取典型单位时已对所研究对象进行了全面分析,故可以用典型调查的结果来精确地推断总体。()
7. 单一调查表就是在一张表上只登记一项调查内容的表格。()
8. 典型调查有“解剖麻雀”和“划类选典”两种。()
9. 如果调查间隔时间相等,这种调查就是经常性调查。()
10. 利用互联网调查只能搜集二手统计资料,不能搜集第一手统计资料。()

二、单项选择题

1. 从一批袋装奶粉中随机抽取 1 000 包进行质量检验,这种调查是()。
- A. 普查 B. 重点调查 C. 抽样调查 D. 典型调查
2. 对家乐福超市全体员工进行身体健康状况调查,调查单位是()。
- A. 每位员工 B. 所有员工 C. 所有商品 D. 每一件商品
3. 普查人口 2010 年 11 月 1 日零时的状况,要求将调查单位的资料在 2010 年 11 月 10 日前登记完成,则普查的标准时间是()。
- A. 2010 年 10 月 31 日 24 时 B. 2010 年 11 月 10 日零时
- C. 2010 年 11 月 9 日 24 时 D. 2010 年 11 月 1 日 24 时
4. 某市 2015 年工业企业经济活动成果的统计年报的呈报时间为 2016 年 1 月 31 日,则调查期限为()。
- A. 一年 B. 一年零一个月
- C. 一个月 D. 一天
5. 调查大庆、胜利、大港、中原的几个大油田,以了解我国石油工业生产的基本情况,这种调查属于()。
- A. 普查 B. 重点调查 C. 抽样调查 D. 典型调查
6. 有意识地选取若干块水田,测算其粮食产量来估算该地区的粮食收成情况,这种调查属于()。
- A. 普查 B. 重点调查 C. 抽样调查 D. 典型调查
7. 下列情况的统计调查,哪一种属于一次性调查()。
- A. 商品库存量 B. 商品购进额
- C. 商品销售量 D. 商品销售额
8. 在下列调查中,调查单位与填报单位一致的是()。
- A. 公司设备调查 B. 农村耕地调查
- C. 学生学习情况调查 D. 汽车养护情况调查
9. 要了解糖果包装的改变对销售量的影响情况,则选定一个地区,将新旧两种包装的糖果投入市场进行试验对比,观察其销售量变化和消费者反应,获得数据作为新包装是否采用的依据。这种调查方法是()。
- A. 观察法 B. 实验法 C. 报告法 D. 访问法
10. 要调查人群中经常上网浏览的人的年龄、性别、职业等情况,比较适宜的调查方法是()。

- A. 观察法 B. 访问法 C. 实验法 D. 网上调查

三、多项选择题

1. 要调查一个地区的学校情况，每一个学校是（ ）。
A. 重点单位 B. 调查单位 C. 调查对象
D. 总体单位 E. 填报单位
2. 下列调查中，属于一次性调查的有（ ）。
A. 人口普查
B. 职工家庭收支变化调查
C. 第三产业从业人数调查
D. VCD 库存量调查
E. 公司利润调查
3. 在工业设备普查中，（ ）。
A. 工业企业是调查对象 B. 工业企业的全部设备是调查对象
C. 每台设备是调查单位 D. 每个工业企业是填报单位
E. 每台设备是填报单位
4. 普查属于（ ）。
A. 全面调查 B. 非全面调查 C. 经常性调查
D. 一次性调查 E. 专门调查
5. 调查单位是（ ）。
A. 需要调查的那些社会经济现象的总体
B. 所要调查的社会经济现象总体的每个单位
C. 调查项目的承担者
D. 负责向上报告调查内容的单位
E. 调查中所调查的具体单位
6. 统计报表的资料来源有（ ）。
A. 原始记录 B. 调查问卷 C. 基层单位内部报表
D. 基层单位统计报表 E. 统计台账
7. 下列调查中属于直接调查搜集第一手资料的方法有（ ）。
A. 观察法 B. 电话调查法 C. 实验法
D. 文献法 E. 互联网调查
8. 要检查节日市场上的肉制品质量，有关部门采用随机原则抽取部分肉制品进行调查，此项调查属于（ ）。
A. 全面调查 B. 非全面调查 C. 典型调查
D. 抽样调查 E. 专门调查
9. 我国第六次人口普查的标准时间是 2010 年 11 月 1 日零时，下列情况应统计人口数的有（ ）。
A. 2010 年 11 月 2 日出生的婴儿

- B. 2010 年 10 月 31 日出生的婴儿
 - C. 2010 年 10 月 31 日晚死亡的人
 - D. 2010 年 11 月 1 日 1 时死亡的人
 - E. 2010 年 10 月 31 日出生，11 月 1 日 5 点死亡的人
10. 网上调查与传统调查相比的优越性在于（ ）。
- A. 费用低
 - B. 无地域、时空限制
 - C. 及时性
 - D. 资料能反映所有用户的信息
 - E. 易获得更完整的统计资料

四、填空题

- 1. 调查表有_____和_____两种形式。
- 2. 专门调查主要有_____、_____、_____和_____ 四种形式。
- 3. 按被研究总体的范围不同来划分，统计调查可分为_____和_____。
- 4. 调查人员亲临现场对调查单位进行清点和计量，这种调查方法称为_____。
- 5. 典型调查中的典型单位是_____选取的，抽样调查中的样本单位是_____选取的。
- 6. 无论采用何种调查方法进行调查都要先制定_____。
- 7. 调查表一般由_____、_____和_____三部分组成。
- 8. 由调查人员通过口头、书面等方式向被调查者了解情况，取得第一手统计资料的调查方法称为_____。
- 9. 调查问卷中问题的设计有两种，即_____问题和_____问题。
- 10. 网上调查分为_____和_____ 两种。

五、应用能力训练题

- 1. 某学院拟对大学生的课外生活进行一次问卷调查。调查项目包括：课外时间的分配和利用、课外活动的形式及时间占用、对大学生勤工俭学的看法和建议。试设计简单的统计调查方案。
- 2. 指出下列调查的调查对象、调查单位和填报单位。

调查内容	调查对象	调查单位	填报单位
手机质量调查			
网络用户年龄调查			
超市商品物价调查			
居民住宅调查			
高等院校调查			
大学生睡眠时间调查			
城市居民生活状况抽样调查			

3. 从统计调查对象包括的范围、调查登记时间是否连续、搜集资料的方法是否相同等方面对以下统计调查实例分类，并指出各属于哪种统计调查方式。

- (1) 2015 年进行全国 1%人口抽样调查。
- (2) 2013 年，对全国所有第二、三产业活动单位进行基本情况摸底调查，以 2013 年 12 月 31 日为标准时点，调查 2013 年度的资料。
- (3) 对进口的一批产品，抽检其中的少部分以对整批产品质量进行评价。
- (4) 要了解全国钢铁产量的基本情况，只要对全国少数几个重点钢铁企业如鞍钢、宝钢、首钢、武钢等进行调查，就能及时对全国钢铁产量的基本情况推断。
- (5) 为了探讨一项新改革措施实施的效果，推广其成功经验，对已采取改革措施并效果明显的代表性单位进行调查。
4. 指出下列调查的统计调查方法，它们属于直接调查还是间接调查。

调查描述	统计调查方法	直接调查或间接调查
(1) 为获得某高速路口汽车交通量资料，调查人员亲自在路口对出入口各种汽车数量进行观察、计数		
(2) 食品厂将新旧两种配方的面包，委托食品店调查各自销售量，以决定是否更改配方		
(3) 经过周密设计，对抽选出的 1 000 位市民发放调查问卷，进行“社会热点问题看法”专题调查		
(4) 为了解同学们使用计算机的习惯，在网上设计问卷进行“校园 E 族调查”		

5. 以你所在的班级为总体进行统计调查，总体单位是每一位同学，调查的有关标志是学生的身高、体重、性别和年龄。不出现姓名。

- (1) 请设计一个简单的统计调查方案。
- (2) 设计一个单一调查表，包括表头、表体和表脚。

每个同学将各自标志的结果写在纸条上，搜集在一起，就形成了原始资料。请保存好这些资料，在后面的章节中我们会用到这些数据。

6. 2015 年 1 月，全国老龄办、民政部、财政部联合下发通知，正式启动第四次中国城乡老年人生活状况抽样调查。第四次调查样本规模大、涉及范围广，参与调查的工作人员将近 4 万人，这样大规模的调查在老龄工作历史上是第一次。除香港特别行政区、澳门特别行政区和台湾省之外，共有 31 个省（自治区、直辖市）及新疆生产建设兵团的 466 个县（区）、1 864 个乡镇（街道）、7 456 个村（居）参与调查工作，首次实现覆盖全国范围的调查目标，其样本量达到 22.368 万，总抽样比约为 1‰。这次调查按老年人口比例抽样，山东、四川、江苏、河南、广东、湖南、安徽、河北、湖北、浙江、辽宁、广西等 12 个省区样本量都在 8 000 份以上，总样本量达 14.9 万份，占全国样本量的 67%。

据悉，第四次调查从筹备启动到全部结束，要持续 3 年左右的时间。2014 年已经做了大量的前期准备工作，包括筹建专家队伍、设计调查问卷等。2015 年是实施调查的关键年，将调查工作细化成了五个阶段：4 月份之前是筹备阶段；4~5 月是动员阶段；5~7 月是组建队伍和培训阶段；8 月是调查阶段；9~12 月是数据处理阶段。调查的标准时点为 2015 年 8 月 1 日零时，

各地于8月1日至31日组织入户调查，并填写社区、乡镇街道、县级组织问卷。

据了解，第四次调查的相关数据、统计结果、专题分析和研究报告，将把调动吸引全社会的广泛关注和促进老龄事业发展顶层设计的完善这两者结合起来，在问卷内容设计上与时俱进，增加了不少新的调查项目，特别是一些带有普遍性、长远性的问题，如老年人口的健康医疗、照料护理服务情况，宜居环境评价，维权意识与行动情况，社会参与状况等。这些调查结果将为老龄产业，特别是养老服务业的改革指明方向。（来源：光明网 2015-04-20）

请根据上述材料信息，利用所学知识，编制尽可能详细的统计调查方案。

7. 你所在的学院伙食如何？大二学生每月零花钱数额为多少，来源及去向是怎样的？大学生就业初职预期薪水和地域如何？对这些问题或你感兴趣的其他问题，进行模拟统计调查。尝试设计一份调查问卷，注意利用调查问卷设计技巧。



本章案例

早做谋划 立足现实 毕业生就业更趋理性 ——2015年本市高校应届毕业生就业情况调查报告之二

据统计，2015年上海高校毕业生有17.7万人，总量压力和结构性矛盾依然存在，就业创业工作形势仍然复杂严峻。近期，国家统计局上海调查总队连续第五年对应届毕业生择业意向展开调查。此次调查采用问卷调查与座谈会走访相结合的方式，调查对象涉及上海交通大学、华东师范大学、上海立信会计学院、华东理工大学、上海东海职业技术学院等16所高校的2071名应届毕业生。受访毕业生中，专科（高职）、本科、研究生（包括硕士、博士生）比例分别约为29.1%、59.5%、11.4%，上海以外地区生源占四成。样本分布情况见下表。

类 别	人数（人）	比例（%）
合计	2 071	100.0
按学历分		
大专	602	29.1
本科	1 233	59.5
研究生	236	11.4
按性别分		
男性	986	47.6
女性	1 085	52.4
按生源类型分		
上海	1 233	59.5
上海以外地区	838	40.5
按专业分		
文科类	1 148	55.4
理工类	760	36.7
其他	163	7.9

续表

类 别	人数 (人)	比例 (%)
按学校类型分		
重点大学	615	29.7
一般大学	915	44.2
大专类院校	541	26.1

注：1. 文科类包括哲学类、经济类、法学类、教育学类、文学类、历史学类、管理学类、艺术类 8 个专业，理工类包括理学类、工学类、农学类、医学类 4 个专业，其他包括军事学类和其他专业。

2. 重点大学指“985 工程”或“211 工程”大学，一般大学指除重点大学以外的高等学校。

一、充分准备、合理定位——择业更趋理性

1. 在校期间为就业做了较充分的准备

调查显示，受访毕业生中毕业前 1 年开始为毕业后出路做准备的占 63%，刚入学就开始做准备的占 11.8%，毕业前 2 年开始做准备的占 11.2%，基本没有准备的占 14%。有 86.5% 的受访毕业生参加过社会活动，有 84.1% 有工作或实习经历，有 67.7% 参加过就业指导，有 26.4% 参加过注册会计师、律师、教师等证书考试。调查同时显示，充分准备是帮助毕业生顺利落实去向的重要因素。有准备的受访毕业生中 64.6% 已有去向，高于无准备者 10.5 个百分点；有过工作或实习经历的受访毕业生中 66.2% 已有去向，高于无经历者 8.6 个百分点；参加过社会活动的受访毕业生中 63.5% 已有去向，比未参加过社会活动的高 2.3 个百分点。此外，有 72.7% 的受访毕业生表示参加就业指导，笔、面试技巧培训等为择业、就业带来了帮助。

2. 选择政府机关和行政事业单位持续降温

政府机关和行政事业单位收入和就业稳定，在当前经济环境下仍是择业首选，调查中有 33.3% 的受访毕业生择业首选政府机关或行政事业单位，其次为“外资企业”（32.1%）、“国有企业”（22.8%）和“私营企业”（9.5%）。但与 2013 年、2014 年对比的调查显示，择业方向出现转变，“公务员热”继续降温，比 2013 年、2014 年分别下降 3.2 个、0.7 个百分点。外企继续受青睐，2015 年选择外资企业的比重达到 32.1%，与选择机关事业单位比重几乎持平，比 2013 年、2014 年分别上升 5.6 个、1.4 个百分点。高校毕业生虽对私营企业依旧缺乏热情，但较 2013 年、2014 年比重分别上升 3.3 个、2 个百分点。自身发展前景、用工权益保障和收入稳定性，仍是阻碍大学毕业生走进私营企业的三大顾虑。

不同群体的毕业生择业方向显现出较大差异。从性别看，38.8% 的女生倾向于政府机关或行政事业单位，高于男生 11.7 个百分点；28.8% 的男生倾向于国有企业，高于女生 11.5 个百分点。从学校类型看，39.2% 的重点大学倾向选择外资企业，高出平均值 7.1 个百分点。

3. “薪酬和福利待遇”是高校毕业生最看重的“硬条件”

对于择业时考虑最多的因素（限选 3 项），在“薪酬和福利待遇”、“解决户籍”、“工作地点”、“发展前景”、“与兴趣、性格相符”、“企业（单位）文化和工作氛围”、“企业（单位）性质与规模”7 项因素中，“薪酬和福利待遇”始终是受访毕业生择业最看重的“硬条件”，有 77.4% 的受访毕业生认为此项因素最重要。同时，发展前景成为毕业生们最看重的“软条件”，有 58.5% 的受访毕业生认为此因素最重要。毕业生综合考虑当前利益和长远发展，更加理性、务实。

择业因素的重要程度对比数据显示，不同择业倾向毕业生侧重的考虑因素有明显差异。倾



向机关事业单位的毕业生有 81.1%选择薪酬和福利待遇，倾向国有企业的毕业生有 65.5%选择发展前景，倾向外资企业的毕业生有 30%选择企业（单位）文化和工作氛围，倾向私营企业的毕业生有 53.1%选择与兴趣、性格相符，以上比重均明显高于倾向其他类型单位的毕业生。此外，对于工作地点，上海生源毕业生较为看重（49.4%），高于非上海生源毕业生 17.9 个百分点；尚未找到工作的受访毕业生选择工作地点的比重（46%）也明显高于其他已落实去向的毕业生（39.8%）。

4. 留沪和返乡成为毕业生最集中的地域选择

调查显示，有 80.2%的受访者希望留沪工作。其中，非上海生源毕业生中，59.1%希望留沪，29.8%希望返乡，11.1%希望到其他城市。选择留沪工作的主要原因（限选 3 项）依次为：环境熟悉（60%）、工作机会多（58.1%）、家庭原因（32.8%）、喜欢上海（25%）、社会关系多（22.8%）、收入高（19.7%）等。选择返乡工作的主要原因（限选 3 项）依次为：环境熟悉（45.2%）、家庭原因（43.6%）、家乡情结（33%）、社会关系多（26.6%）、生活压力小（26.1%）、发展前景好（25%）。由此可见，毕业生无论是选择留沪还是返乡，环境熟悉和家庭原因都是非常重要的影响因素。问及“选择就业单位和就业地时，会充分考虑哪些方面的意见（限选 3 项）”，有 73.4%的毕业生选择“父母的意见”，说明父母的想法对毕业生就业、择业起到至关重要的作用。

二、期待较高、困惑不少——就业需更务实

1. 现实与理想差距较大，近四成签约民营或私营企业

调查显示，与毕业生倾向机关事业单位和外资企业的择业预期不同，在已签约的近 700 名受访毕业生中，36.5%的签约单位为私营企业，29.4%为国有企业，19.1%为外资企业，11.6%为政府机关或行政事业单位，3.5%为其他。私营企业已经成为吸纳毕业生就业的主要力量。

2. 实际签约薪酬和期望值相差逾千元

据已签约毕业生薪酬数据计算，2015 年本市应届毕业生实际签约的税前月薪平均约为 4 800 元，比 2014 年调查结果高 400 元左右。从学历看，专科（高职）平均税前月薪约为 3 500 元，本科生为 5 100 元，研究生为 6 800 元。另一方面，据调查数据加权计算，2015 年应届毕业生期望的税前月薪平均为 5 800 元，比 2014 年调查结果高约 500 元。总体来看，毕业生期望薪酬与实际获得水平相差逾千元。

调查显示，71.7%的已签约受访毕业生对签约结果表示满意，对工作稳定、工作地点、单位文化和工作氛围、单位性质等方面满意度相对较高。然而，对毕业生就职最看重的因素——薪酬和福利待遇，满意度却相对较低，63.2%表示满意，32.1%表示一般，4.7%表示不满意。

3. 逾三分之一受访毕业生无清晰的职业目标

清晰的职业目标是毕业生发掘自我潜能、增强个人实力的重要激励因素，也是能否顺利就业的重要条件。调查显示，65.2%的受访毕业生表示有清晰的职业目标，34.8%表示没有。有清晰职业目标的受访毕业生中有 66.2%有去向，高于无目标者 8.6 个百分点。据调查数据加权计算，有清晰职业目标的受访毕业生签约税前月薪为 4 900 元，高于无目标者约 400 元。

三、对应届毕业生求职就业的几点建议

1. 首次就业未必终身职业，应立足当下、规划长远

择业是高校毕业生由学生转变为就业者的第一步。有调查显示，2010 届大学生在毕业三年内有 41%转换了岗位，有 48%转换了行业。可见，应届毕业生毕业后第一份工作不一定是终身

职业，首次求职很难一步到位、一劳永逸，建议应届毕业生要立足当下、长远规划，先就业再择业。在实际工作中不断学习和积累经验，找准自身定位，寻找发展路径，向自己理想的行业和职位努力。

2. 打有准备之战是缩小期待与现实差距的重要途径

调查显示，已签约的受访毕业生认为求职成功最关键的因素（限选3项）排在前三位的依次为：专业背景（44.1%）、有工作或实习经历（41.8%）、对收入、待遇等要求合理（20.9%）。在签约月薪为6 000元以上的受访毕业生中，有50.6%、24.1%和20.5%认为是有工作或实习经历，笔、面试技巧和在校期间的社会活动经历，分别较平均值高出8.8个、8个和7.7个百分点。可见，扎实的专业基础、较丰富的实践经历是毕业生顺利就业不可或缺的基本条件。建议大学生在校期间就为毕业后的出路早做准备和规划，结合理想职业的实际需求而努力，打有准备之战，以缩小期望和现实的差距。

3. 放开视野，为职业发展争取更多机会和可能

调查中，尚有相当一部分毕业生未落实工作。分析原因，这部分毕业生除了准备不充分、自身硬性条件不够外，过分看重就业地域和行业也是一个重要的影响因素。为此，建议毕业生不要故步自封，要放开视野，对就业的地域、行业作更广泛的选择，为自身的职业发展争取更多机会和可能。在政府支持、社会关心、家长费心等大环境下，自立自强、自主择业，克服等机会、靠父母等消极思想和依赖思维，以就业这个人生重大转折点为契机，真正实现从学生到社会人的转变。

（来源：上海市统计局 2015-06-23）

统计整理

学习要点

- 统计整理的含义、内容及步骤。
- 统计分组的基本理论和方法。
- 分配数列的特性与编制方法。
- 编制统计表、绘制统计图。

在对实际问题的研究过程中，需要大量的数据，可以通过统计调查搜集数据。但是，统计调查获得的是原始资料，这些数据只能反映出总体各单位的具体情况。例如，民意测验收集到每张调查者填写的调查表格、每个顾客反馈给企业的调查表格、中国互联网络信息中心（CNNIC）进行的中国互联网发展情况的调查所得的原始资料等，都是无序的、杂乱的、无规律的。为了充分利用搜集到的统计信息进行统计分析，必须将杂乱无章的资料理顺。例如，为了了解市民对某项政策的态度，就必须对民意调查表格进行整理，将“满意”、“一般”、“反对”的人数分别统计，进行比较得出结论；企业只有将顾客调查表格作以整理分析，才能知道消费者的知识结构、收入水平、消费倾向，进而制订企业的生产、销售等计划；中国互联网络信息中心必须把原始数据汇总后才能形成上网计算机数、用户人数、用户分布、信息流分布等统计信息，才能为国家和企业了解互联网发展情况提供依据并据此做出决策。所以，统计整理是统计工作中不可缺少的重要环节，本章将对其加以介绍。

3.1 统计整理的意义和内容

3.1.1 统计整理的意义

统计整理（statistical treatment）指根据统计研究的任务和要求，对统计调查所取得的原始资料进行科学的分类、汇总或对已整理过的资料进行再加工，使之系统化、条理化，以得出反映现象总体特征的综合资料的工作过程。

通过统计调查所搜集的数据资料是杂乱无序的，仅仅反映了总体单位的具体情况和事物的表面现象，不能深刻说明事物的本质，难以揭示出事物的内在发展规律。因此，必须对这些资料进行科学的加工整理，以便概括说明总体的特征和认识其内在规律性。

统计整理是统计工作中不可或缺的重要环节，具有承前启后的作用。统计整理不仅是一个简单的汇总工作，它既是统计调查的深入和继续，也是统计分析的基础和前提。如果没有统计整理，即使统计调查所得的资料再丰富、再完善，其作用也发挥不出来，统计工作也就无法进行下去。统计调查所搜集到的大量资料，只有通过科学的整理加工，形成反映总体特征的综合资料，才能为下一步的统计分析工作做好准备。统计整理的结果是否如实反映客观情况，决定了统计资料的价值，进而影响到统计分析的准确性与真实性。

3.1.2 统计整理的内容

统计整理工作是一项细致的、科学的工作，统计整理包括对统计资料的审核、分组、汇总和编制统计图表等环节，因此，进行统计资料整理的主要内容包括以下几个方面。

1. 拟订统计整理纲要

在进行统计整理之前，需要根据目的，确定对统计调查所搜集到的资料的哪些内容进行整理、对统计资料如何分组，以及确定整理中所涉及的指标和指标体系。因此，统计整理的首要工作就是拟订整理纲要。整理纲要的主要形式是一整套整理表格，在表格上列出指标体系和分组，另外，还要编制相关说明。整理纲要是保证统计整理工作有计划、有步骤、有组织进行的前提。

2. 审核原始资料

在搜集资料的过程中，经常会由于某些原因出现一些差错。因此，为了确保统计资料准确无误及符合统计研究目的的要求，必须对统计调查所获得的原始资料进行严格的审核，发现问题及时纠正。它是统计整理中的一个重要的环节，主要包括资料的准确性、及时性和完整性审核。

审核资料的准确性，就是检查所填报的资料是否准确可靠。它主要从两个方面进行：其一，逻辑检查，主要从理论上或常识上检查资料的内容是否有悖常理、有无不切实际或不符合逻辑的地方及各项目之间有无相互矛盾之处；其二，计算检查，主要检查填报单位有无遗漏、调查表项目是否填齐、所填内容和表格规定是否一致、计量单位与法定单位是否一致，以及各项数字之间的关系是否正确等。

审核资料的及时性，主要检查资料是否符合调查规定的时间，是否在规定的调查期限报出。

审核资料的完整性，主要检查调查单位资料是否齐全，是否上报了规定的份数、调查项目是否完整等。

3. 进行分组汇总

分组汇总就是用一定的组织形式和方法对经过审核的资料进行分组、汇总和计算。根据研究目的和统计分析的需要，选择整理的标志，并进行划类分组。科学的分组是搞好统计整理的前提条件，只有正确地分组才能整理出有科学价值的综合指标，并借助这些指标来揭示现象的本质与规律。

4. 编制和绘制统计图表

根据统计分析的要求或社会经济现象之间的联系，对整理好的资料编制成统计表或绘制成统计图，以简明扼要地反映社会经济现象之间的数量特征。

5. 统计资料的保管与积累

统计研究中经常要进行动态分析，这就需要有长期累积的历史资料，而根据积累资料的要求，对已有的统计资料进行筛选，以及按历史的口径对现有的统计资料重新调整、分类和汇总等，都必须通过统计整理工作来完成。

3.2 统计分组

3.2.1 统计分组的概念和作用

统计分组（statistical classification）就是根据统计研究的需要和总体的内在特征，将总体按照一定的标志划分为若干个组成部分的一种统计方法。简而言之，统计分组就是把性质相同的总体单位归为一组，不同性质的总体单位区别开来形成不同的组。社会现象是复杂的，现象之间既有某种共同的性质，同时又存在差异，通过统计分组，揭示现象各部分之间存在的差异，结合统计分析，能够认识事物的发展变化规律。例如，企业按资产大小划分为大型企业、中型企业和小型企业；我国将国民经济产业结构划分为第一产业、第二产业和第三产业；考试成绩经常划分为60分以下、60~70分、70~80分、80~90分、90分及以上五个等级等。统计分组的结果是保持组内资料的一致性和组间资料的差异性。

理解统计分组应从两个方面进行：一方面，对总体而言是“分”，即将总体按某个标志划分为若干个性质相异的组成部分；另一方面，对总体单位而言是“合”，即将具有某种相同性质的总体单位归在同一组。

统计分组是统计整理的关键，只有对总体进行科学的分组，才能对统计资料进行进一步的科学整理和分析，以得到有价值的结论。

统计分组是统计研究中应用的基本方法之一，贯穿于统计研究的整个过程，在统计研究中占有非常重要的地位。统计分组的作用主要体现在以下三个方面。

1. 划分现象的类型

划分现象的类型是统计分组的主要作用。社会经济现象千差万别，存在着不同的类型，而且各种不同的类型具有不同的特点及不同的发展规律，要了解各种社会经济现象的性质、特点、相互关系及变化规律，必须运用统计分组将现象总体划分为性质不同的类型，以便揭示不同社会经济现象的差异。例如，根据登记注册类型划分的我国城镇单位就业人员平均工资的情况，如表3.1所示。

表 3.1 2013 年按登记注册类型划分的城镇单位就业人员平均工资

	城镇单位就业人员平均工资（元）
全国	51 483
国有单位	52 657
城镇集体单位	38 905
股份合作单位	48 657
联营单位	43 973
有限责任公司	46 718
股份有限公司	61 145
其他内资	38 306
港、澳、台商投资单位	49 961
外商投资单位	63 171

（资料来源：中国统计年鉴 2014）

通过这种分类，可以反映我国城镇就业人员在不同类型的企业就业的情况，也为进一步研究我国城镇人员就业结构提供了依据。

2. 反映现象总体的内部结构

从数量上反映总体的内部结构是统计研究的重要任务。社会现象包含大量的单位，它们在性质上不尽相同，而且在各种类型总体中的比重也不同，因而所处的地位和对总体的影响程度也不同。不同的类型分布决定了不同的总体性质。将社会经济现象总体经过统计分组后划分为不同性质的组成部分，计算总体内部各组成部分占总体的比重，分析总体各部分的性质、结构和比例关系。另外，还可以将总体内部结构分组资料按时间的发展顺序进行对比分析，从而认识现象发展变化的规律。例如，2009—2014 年我国按城乡分类的就业人员数（年底数），如表 3.2 所示。

表 3.2 2009—2014 年我国按城乡分类的就业人员数（年底数）

单位：万人

年 份	城 镇	乡 村	合 计
2009	33 322	42 506	75 828
2010	34 687	41 418	76 105
2011	35 914	40 506	76 420
2012	37 102	39 602	76 704
2013	38 240	38 737	76 977
2014	39 310	37 943	77 253

（资料来源：中国统计年鉴 2014，2014 年国民经济和社会发展统计公报）

从表中可以看出，2009—2014 年 6 年间我国城乡中的就业人员发生了变化，城镇就业人员数逐渐上升，乡村就业人员数逐渐下降。

3. 分析现象之间的依存关系

统计研究的社会经济现象不是彼此孤立的，总是存在相互依存和相互制约的关系。例如，

国民收入与居民储蓄额、市场商品价格与其需求量之间、企业销售额和流通费用率、家庭的工资收入与生活费支出之间、工人技术级别与产品质量之间、工人劳动生产率与产品成本之间，等等，都在一定程度上存在相互依存的关系。这些依存关系，可通过统计分组来发现影响因素与结果因素之间的变动规律，并可进一步从数量上描述依存关系的密切程度。这些内容将在第9章中详细阐述。表3.3可以反映我国人均国内生产总值与城镇居民人均可支配收入的关系。

表 3.3 我国人均国内生产总值与城镇居民人均可支配收入的关系

年 份	人均国内生产总值（元）	城镇居民人均可支配收入（元）
2010	30 015	19 109.4
2011	35 198	21 809.8
2012	38 459	24 564.7
2013	41 908	26 955.1
2014	46 652	28 844.0

（资料来源：中国统计年鉴 2014，2014 年国民经济和社会发展统计公报）

从表 3.3 中可以看出，人均国内生产总值数值在不断变化，其对应的城镇居民人均可支配收入也在变动，这说明我国人均国内生产总值的不断增加，导致城镇居民人均可支配收入的持续增长，两者之间存在相互依存关系。

3.2.2 统计分组的种类

1. 按分组的作用和目的划分

按分组的作用或目的不同，统计分组分为类型分组、结构分组和分析分组。

这种分组的种类主要是从统计分组的基本作用的角度出发的。由于统计分组的作用体现在三个方面，因此，分组的类型也相应划分为类型分组、结构分组与分析分组。

- （1）类型分组是将复杂的现象总体划分为若干个不同性质的部分。
- （2）结构分组是在对总体分组的基础上计算出各组对总体的比重，以此来研究总体各部分的结构。
- （3）分析分组是为研究现象之间的依存关系而进行的统计分组。

一般情况下，类型分组和结构分组总是紧密联系在一起。

2. 按分组标志的多少划分

按分组标志的多少及其排列形式，统计分组可分为简单分组、复合分组和分组体系。

这种分组类型主要是从分组时所选用的分组标志的数量多少的角度出发的。统计分组时可以有单一标志作为分组标志，也可以同时采用多个标志作为分组标志。

（1）简单分组也称为单一分组，就是对总体只按一个标志进行分组。例如，上网计算机按上网方式划分为拨号上网、专线上网、ISDN 上网、宽带上网和其他方式上网等；企业按生产规

模可分为大型、中型和小型三组；货运按运输方式可分为铁路运输、公路运输、水陆运输、航空运输与管道运输五组。

（2）复合分组就是对所研究的总体按两个或两个以上的标志进行的多层次分组，即在按某一标志分组的基础上，再按另一标志进一步分组。例如，人口按性别分组后再按年龄分组；企业按经济类型分组后再按规模进行分组。再如，旅客先按运输量不同计算形式分组，然后再按不同运输方式分组，如表 3.4 所示。

表 3.4 2014 年我国各种运输方式完成旅客运输量及其增长速度

指 标	单 位	绝 对 数	比上年增长 (%)
旅客运输总量	亿人次	220.7	3.9
铁路	亿人次	23.6	11.9
公路	亿人次	190.5	2.8
水运	亿人次	2.6	12.3
民航	亿人次	3.9	10.6
旅客运输周转量	亿人公里	29 994.2	8.8
铁路	亿人公里	11 604.8	9.5
公路	亿人公里	11 981.7	6.5
水运	亿人公里	74.4	8.9
民航	亿人公里	6 333.3	12.0

（资料来源：2014 年国民经济和社会发展统计公报）

采用复合分组方式，可以从同一现象的层层分组和分组标志的联系中更深入地反映总体的内部结构，更细致全面地分析问题。当采用一个分组标志不能充分说明现象或现象受两个或两个以上因素影响时，选用复合分组比简单分组更能清晰地反映现象的结构。但是，复合分组的组数会随着分组标志的增加而成倍增加，使每组包括的单位数相应减少，处理不好就会影响分析问题。因此，在采用复合分组时，分组标志选择要适量，而且总体包括的单位数应较多。究竟采用几个分组标志进行分组，要根据统计研究的目的和任务来决定。

（3）分组体系就是采用一系列相互联系、相互补充的并列标志对被研究对象总体进行的分组。社会现象是复杂的，无论简单分组还是复合分组，都只能对现象从一个或几个方面进行观察与分析，因此，为获得对复杂现象全面的认识，就应采用分组体系对某一经济现象从不同角度、不同方面进行观察和分析研究。例如，研究我国网民状况时，可以按性别、婚姻状况、年龄、受教育程度、个人月收入、职业、行业等多种标志进行分组，这些标志互相联系、互相补充，充分说明了我国网民的构成状况。

3. 按分组标志的性质划分

按分组标志的性质，可分为品质标志分组和数量标志分组。

（1）品质标志分组就是按品质标志进行的分组，即按事物的某种属性分组。品质标志分组可以反映总体内部的性质差异，如企业按经济类型分组，人口按性别、民族、受教育程度、职业分组等。



（2）数量标志分组就是按数量标志进行的分组，如人口按年龄分组、企业按固定资产价值分组、网民按个人月收入分组等。数量标志分组可以反映现象总体数量上的差异。

3.2.3 分组标志选择及界限的确定

统计整理的关键在于统计分组，统计分组的关键在于正确选择分组标志和划分各组界限。分组标志是统计分组的依据或标准，正确地选择分组标志能使统计分组的作用得以充分发挥。

反映总体单位的标志有许多，标志的选择将影响统计分析的结论的正确性与真实性。正确选择分组标志，必须遵循以下几条原则。

1. 应根据研究的目的与任务选择分组标志

统计研究的目的是统计分组标志选择的依据。统计分组是为统计研究服务的，统计研究的目的不同，对分组的要求也不同，选择的分组标志也应有所不同。例如，要分析我国网民所在行业的结构分布，就要选择行业为分组标志；要分析网民的教育结构，则要选择受教育程度作为分组标志；要分析性别结构，则要选择性别作为分组标志；要分析网民的收入结构，则要选择网民个人月收入作为分组标志。

2. 选择最能体现现象本质特征的标志作为分组标志

研究某一问题可能涉及许多标志，有些是主要标志，有些是次要标志。主要标志能反映现象的本质，因此，在统计分组时，要选择最能体现现象本质特征的主要标志作为分组标志。例如，研究企业的经济效益时，可供选择的分组标志有：总产值、净产值、销售收入、利税额、单位产品成本等，而其中最能反映企业经济效益好坏的标志是利税额，故应选择这一标志作为分组标志。

3. 结合现象发展的具体历史条件和经济条件选择分组标志

社会经济现象随着时间、地点、条件的变化而不断变化，反映现象本质特征的主要标志也会因时、因地而不同。因此，选择分组标志时，根据历史的发展，具体情况具体分析，“与时俱进”，客观选择分组标志。例如，研究工业企业的生产能力，在机械化程度低下的情况下，应选择职工人数作为主要分组标志；在机械化程度高的情况下，应选择固定资产作为主要分组标志。再如，研究企业规模，对于劳动密集型企业，应选择职工人数作为主要分组标志；对于资金密集型企业，则应选择固定资产价值或技术拥有量作为分组标志。

确定分组标志后，还必须划分各组界限。划分各组界限就是在分组标志的变异范围内，划定各相邻组间的性质界限和数量界限。划分各组界限一定要本着保证各组组内单位的同质性和组与组之间单位的差异性的原则进行。

总之，分组标志的选择与分组界限的划分是统计研究中非常关键的一步，必须遵守相关的原则认真分析，做出正确的判断，选出合理的分组标志，为进一步的统计分析做好准备。

3.2.4 统计分组的方法

按分组标志的性质不同，分为品质标志分组和数量标志分组，这种分组类型是最重要的统计分组类型。这两种分组的具体处理方法不同，统计分组方法主要是围绕这两种分组进行阐述的。

1. 品质标志分组的方法

品质标志分组是指选择反映事物属性差异的品质标志作为分组标志后，根据其变异范围划定各组界限，将总体划分为若干个形式不同的组成部分。

按品质标志分组时，有些情况比较简单，有些情况比较复杂。具体说来，其组数的确定主要取决于两个因素——事物的特点与统计研究的任务。事物本身所具有的既定属性，是确定组数的基本依据。进行品质分组，其组数的多少首先取决于事物本身的特点。例如，人口按性别分组，就只能分为两组；高等学校在校生按学历可分为专科、学士、硕士和博士四组。

对于有些事物构成比较复杂，组数可多可少的情况，就需要根据统计研究的目的和要求来确定组数。例如，就业人员按行业进行分组，组数可多可少，究竟分为几组合适，只能根据统计研究的目的和要求来确定了。要求详细时，组数可多些；要求粗略时，组数则可少些。

在统计实践工作中，常常需要对所研究的现象进行复杂的品质分组。对于比较复杂的品质标志分组，为了避免认识不同而造成的混乱，保证分组的同一性和可比性，国家统计局制定了统一的分类目录和标准，作为分组的标准和依据，如《国民经济行业分类》、《关于统计上公有和非公有控股经济的分类办法》、《关于划分企业登记注册类型的规定》等，各地区、各部门进行统计资料整理时必须遵照执行。

2. 数量标志分组的方法

数量标志分组是指选择反映事物数量差异的数量标志作为分组标志后，根据其变异范围划定各组界限，将总体划分为若干个性质不同的组成部分。

数量标志分组，不是简单地确定各组间的数量差异，而是要通过分组体现数量变化来确定现象的不同性质和类型。因此，根据变量值的大小来准确划分性质不同的各组界限时，首先要分析总体中有多少种性质不同的组成部分，然后再确定各组成部分的数量界限。

按数量标志分组的方法有两种：单项式分组和组距式分组。

(1) 单项式分组。单项式分组就是按每个变量值对现象总体进行的分组，如某区居民家庭拥有汽车数，如表 3.5 所示。

表 3.5 某区居民家庭拥有汽车数情况表

居民家庭汽车拥有数量（辆）	居民户数（户）
0	68
1	24
2	6
3 以上	2
合 计	100

单项式分组适用于离散变量并且变量值的变动范围不大的情况。

（2）组距式分组。组距式分组就是按变量值的一定范围对现象总体进行的分组。将现象的总体变动范围划分为若干个区间，以每个区间作为变量值。例如，我国网民个人月收入构成 2015 年 6 月与 2014 年 12 月的对比情况，如表 3.6 所示。

表 3.6 我国网民个人月收入构成 2015 年 6 月与 2014 年 12 月的对比情况

按收入分组（元）	2015 年 6 月网民个人月收入比重（%）	2014 年 12 月网民个人月收入比重（%）
无收入	6.2	2.2
1~500 元	11.2	15.2
500~1 000 元	9.8	10.7
1 000~1 500 元	7.3	8.1
1 500~2 000 元	8.9	10.3
2 000~3 000 元	21.0	18.8
3 000~5 000 元	22.4	20.2
5 000~8 000 元	8.2	8.7
8 000 元以上	5.0	5.7
合 计	100.0	100.0

（资料来源：中国互联网络信息中心 第 36 次中国互联网络发展状况统计报告）

组距式分组适用于连续变量和变动范围较大的离散变量。

3.3 分配数列

3.3.1 分配数列的意义和种类

在统计分组的基础上，将总体的所有单位按组进行归类整理并按一定顺序排列，计算出各组的单位数，形成一个反映总体中各单位在各组中的分布情况的数列，这个数列称为分配数列或次数分布（frequency distribution）。

分配数列是统计整理的一种重要形式，也是进行统计描述和统计分析的基础。它可以反映总体的结构分布状况和分布特征，并且由此可以分析研究总体中某一标志的平均水平及其变动规律。

分配数列由两要素构成：一是总体中按某标志分的组，二是各组相应的分配次数或频率或标志值。在分配数列中，分布在各组的总体单位数称为次数，又称为频数（frequency）；各组次数占总体次数的比重，称为比率，又称为频率。次数和频率从不同角度反映各组标志值出现的频繁程度，说明总体单位在各组中的分布情况，是分配数列的两种表现形式。

分配数列按分组标志特征的不同，分为品质分配数列和变量分配数列。

1. 品质分配数列

品质分配数列，又称品质数列，是按品质标志分组形成的分配数列。它在品质分组的基础上形成，由各组名称和次数构成，主要用于研究总体构成情况。各组次数以绝对数表示，可称为频数；以相对数表示，可称为频率，如表 3.7 所示。

表 3.7 我国 2014 年年末城乡人口构成表

按城乡分组	人数（万人）	比重（%）
城镇	74 916	54.77
乡村	61 866	45.23
合计	136 782	100.00

（各组名称）

（次数或频数）

（比率频率）

从表 3.7 中可以看出，该品质数列反映了我国 2014 年年末城乡人口特征：城镇人口多于乡村人口。只要品质标志分组正确，形成的分配数列通常都能准确地反映总体的分布特征。

2. 变量分配数列

变量分配数列，又称变量数列，是按数量标志分组形成的分配数列。它是在变量分组的基础上形成的分配数列，由各组变量值和次数构成，主要用于反映不同变量值各组的分布情况。变量数列可以分为单项式变量数列和组距式变量数列。

（1）单项式变量数列。简称单项数列，是按单项式分组形成的变量数列，每个变量值是一个组，顺序排列。在单项式变量数列中，有多少不重复的变量值就有多少组。单项式变量数列仅适用于变动幅度比较小、不重复变量值较少的离散变量分组的情况。表 3.5 为单项式变量数列。

（2）组距式变量数列。简称组距数列，是由组距式分组形成的变量数列，每个组由若干个变量值形成的区间表示。组距式变量数列适用于连续变量分组或变动幅度较大、不重复值较多的离散变量分组的情况。表 3.6 为组距式变量数列。

3.3.2 变量数列的编制

1. 变量数列次数分布表的编制

变量数列次数分布表的编制也分为单项式变量数列和组距式变量数列两种情况。

对于总体单位数不多的离散变量，可以直接编制单项式变量数列次数分布表。具体编制方法：按变量值分组列于左方，然后将各变量值出现的次数列于右方，如表 3.5 所示。

而对于连续变量或变动幅度较大、不重复值较多的离散变量则要编制组距式变量数列。组距式变量数列次数分布表编制比较复杂，下面就其编制方法作以详细讲解。

1) 原始资料按数值大小依次排序

【实例 3.1】 某高校经济管理系 2015 年某班 40 名学生英语成绩，资料如下：

69	61	75	60	80	72	82	86	90	78
62	87	71	77	79	87	85	97	85	63
94	78	65	85	96	88	89	82	87	78
95	83	56	81	58	75	78	74	67	75

将上述资料进行排序，结果如下：

56	58	60	61	62	63	65	67	69	71
72	74	75	75	75	77	78	78	78	78
79	80	81	82	82	83	85	85	85	86
87	87	87	88	89	90	94	95	96	97

2) 确定组距、组数和组限

编制组距式变量数列时，必须要确定组距和组数。首先应对标志值的分布情况进行仔细观察，找出最大值与最小值并计算出它们的距离（即全距），确定大多数标量值分布在什么范围，然后确定组距和组数。

组距（class interval）是每个组上限与下限之间的距离，组距=上限-下限。组限是组与组之间的界限。每个组的最大值称为上限（upper limit），每个组的最小值称为下限（lower limit）。将上、下限都齐全的组称为闭口组；将有上限缺下限或有下限缺上限的组称为开口组。开口组一般使用“以上”或“以下”来表示。开口组组距以相邻组组距作为参照。

组距与组数关系密切，同一现象的变量分组，组距越大则组数越少，组距越小则组数越多。确定组距与组数应遵循下列原则：尽可能反映总体单位的分布情况及总体单位的集中趋势，体现组内资料的同质性和组与组之间资料的差异性。

组距式变量数列根据组距是否相等，分为等距变量数列和异距变量数列。对于等距变量数列的编制，一般依据总体内部情况的定性分析确定组数，然后用全距除以组数来确定组距，并由此划分各组界限。设 R 为全距， K 为组数， i 为等距组，则 $i = \frac{R}{K}$ 。

在【实例 3.1】中，最小值为 56，最大值为 97，全距=97-56=41。根据考试成绩的性质，将组距确定为 10，组数为 5。

确定组数和组距之后，还要确定组限。组限的确定要遵循这一原则：按这样的组限分组后，标志值在各组的变动能反映总体单位分布的规律性。确定组限时需注意：特殊的界限点必须作为组限（如反映水温的 0°C 和 100°C、反映计划完成程度的 100%、反映学习成绩的 60 分和 100 分等）。最小组的下限要略低于最小变量值，最大组的上限要略高于最大变量值。

连续变量分组，相邻两组的上限与下限通常是相同的，每一组的上限同时是下一组的下限。因此，这种分组又称为同限分组。为了避免计算时总体单位数出现错误，统计各组单位数时应

将到达上限值的单位数计入下一组内，即遵循所谓的“上组限不在内”的原则。

离散变量分组，相邻两组的上限与下限可以相同，也可以不相同。相邻两组的上限与下限相同时也遵循“上组限不在内”的原则。相邻两组的上限与下限不相同，又称为异限分组。例如，工业企业按职工人数分组，如表 3.8 所示。

表 3.8 离散变量分组的两种表示方法

相邻两组上限与下限相同（人）	相邻两组上限与下限不同（人）
100 人以下	99 人以下
100～500	100～499
500～1 000	500～999
1 000 以上	1 000 以上

3) 编制组距式变量数列次数分布表

确定了组距、组数和组限，经过整理就可绘制组距式变量数列表，如表 3.9 所示。

表 3.9 40 名学生英语成绩分布表

按考试成绩分组	人 数	比重（%）
50～60	2	5.0
60～70	7	17.5
70～80	12	30
80～90	14	35
90～100	5	12.5
合计	40	100.0

组距变量数列是以变量的区间进行分组的，这掩盖了各组总体单位的实际变量值。因此，为了反映每个组的一般水平，我们一般采用组中值表示。组中值（mid-value of class）是各组变量的中间数值。组中值的计算公式分开口组与闭口组两类：闭口组的组中值包括同限分组组中值和异限分组组中值两种；开口组的组中值包括缺上限的开口组组中值和缺下限的开口组组中值两个。

同限分组组中值的计算公式为：

组中值 = (上限 + 下限) / 2

异限分组组中值的计算公式为：

组中值 = (本组下限 + 后一组下限) / 2

缺上限的开口组组中值计算公式为：

组中值 = 下限 + (邻组组距) / 2

缺下限的开口组组中值计算公式为：

$$\text{组中值} = \text{上限} - \frac{\text{邻组组距}}{2}$$

【实例 3.2】 计算表 3.10 中各组数据的组中值。

表 3.10 某高校大四学生每月消费水平组中值计算结果表

按每月消费金额分组（元）	人 数	组 中 值
1 000 以下	80	900
1 000～1 200	160	1 100
1 200～1 500	270	1 350
1 500～1 800	240	1 650
1 800～2 000	140	1 900
2 000 以上	110	2 100
合 计	1 000	—

第一组的组中值为：

$$1\,000 - \frac{1\,200 - 1\,000}{2} = 900$$

第二组的组中值为：

$$\frac{1\,200 + 1\,000}{2} = 1\,100$$

第三、四、五组的组中值计算类似于第二组的计算，结果如表 3.10 所示。

最后一组的组中值为：

$$2\,000 + \frac{2\,000 - 1\,800}{2} = 2\,100$$

2. 变量数列次数分布图的绘制

一般情况下，我们不仅可以用次数分布表表示总体单位的分布状况，还可以通过绘制变量数列次数分布图直观形象地揭示总体单位的分布状况和规律性。

单项式变量数列次数分布图的绘制比较简单，具体绘制方法为：直接以变量值为横轴，以次数为纵轴，在坐标轴上描绘出各组变量值与相对应的次数对应的坐标点，连接各坐标点即可得分布图。

组距式变量数列次数分布图一般有直方图（histogram）和次数分配曲线图，但是，次数分配曲线图也是在直方图的基础上绘制的。绘制组距式变量数列次数分布图时需注意等距变量数列与异距变量数列的处理方法不同。等距变量数列的各组单位数只受变量变化的影响，各组单位数可以直接比较，分组后直接统计整理即可绘制等距变量数列次数分布图。例如，利用表 3.9 的资料绘制等距变量数列次数分布图，如图 3.1 所示。

绘制等距变量数列次数分布图的具体步骤如下：以横轴代表变量值，在横轴上标出各组组限值所在位置，以纵轴代表次数并在上面标出各组次数所在位置；然后以组距为宽，以次数为高，绘制相对应的矩形，各组矩形组合构成变量数列总体次数分布特征的直方图。若需要绘制次数分配图，只要在直方图的基础上，连接各矩形顶边的中点（即各组的组中值），形成一条折线，并且应在直方图的左右两端假设各有一个组，将折线与两个假设组的中点连接，就形成了次数分配曲线。这时曲线所覆盖的全部面积与直方图的面积相等。

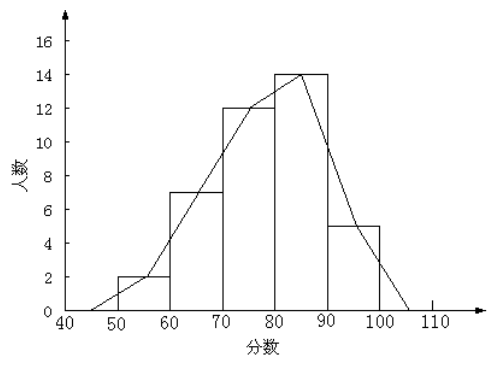


图 3.1 学生英语成绩次数分布图

异距变量数列的各组单位数受到组距大小不等和变量值两种因素的影响，因此需要计算次数密度，以消除不等组距对各组次数的影响，进而准确地反映实际的次数分布。次数密度也称为频数密度，就是单位组距内分布的次数。其计算公式如下：

$$\text{次数密度} = \frac{\text{各组次数}}{\text{各组组距}}$$

通过计算次数密度，将不等组距的次数换算为标准组距次数。某组标准组距次数等于该组标准组距与该组次数密度的乘积。标准组距为异距变量数列中的各组组距中最小的组距。标准组距次数的计算公式如下：

$$\text{某组标准组距次数} = \text{该组次数密度} \times \text{标准组距}$$

绘制异距变量数列次数分布图的具体步骤与等距数列类似，不同的处理地方在于：以纵轴代表各组标准组距次数并在上面标出各组标准组距次数所在位置；然后以组距为宽，以标准组距次数为高，绘制相对应的矩形，各组矩形组合构成直方图。

【实例 3.3】 根据某公司员工每月消费水平绘制变量数列次数分布表和次数分布曲线，如表 3.11 和图 3.2 所示。

表 3.11 某公司员工每月消费水平次数分布表

按每月消费金额分组（元）	人 数	组 距	次数密度	标准组距人数
1 800～2 000	80	200	0.4	80
2 000～2 200	160	200	0.8	160
2 200～2 500	270	300	0.9	180
2 500～2 800	240	300	0.8	160

续表

按每月消费金额分组（元）	人 数	组 距	次数密度	标准组距人数
2 800~3 000	140	200	0.7	140
3 000~3 500	110	500	0.22	44
合 计	1 000	—	—	—

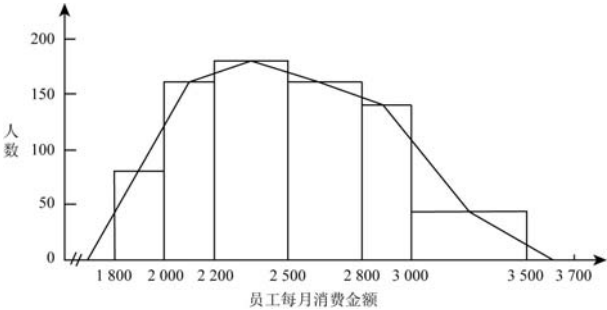


图 3.2 某公司员工每月消费水平次数分布图

3. 累计次数分布

借助于次数分布表和分布图可以观察分配数列中各组的次数及总体单位数的分布特征，但若获取截至某一组变量值以上或以下的分布次数或事物发展进程等情况，就需要编制累计次数分布。如在表 3.9 中，要计算每个班不及格的学生有多少、大于 80 分的学生有多少等数据。

编制累计次数分布表或绘制累计次数分布曲线都需要计算累计次数或累计频率。累计次数与累计频率分别表明总体的某一标志值在某一水平上下的总体次数与比率。累计次数的计算方法有两种：向上累计和向下累计。

（1）向上累计。也称为较小制累计，是将各组次数和比率从变量值低的组向变量值高的组逐组累计。组距数列向上累计的次数表明各组上限以下总共包含的总体次数和比率是多少。

（2）向下累计。也称为较大制累计，是将各组次数和比率从变量值高的组向变量值低的组逐组累计。组距数列向下累计的次数表明各组下限以上总共包含的总体次数和比率是多少。

例如，根据表 3.9 中学生英语成绩次数分布表计算的累计次数分布表如表 3.12 所示，绘制的累计次数分布图如图 3.3 所示。

表 3.12 40 名学生英语成绩累计次数分布表

按考试成绩分组	成绩分布		向上累积		向下累积	
	人 数	比重（%）	频 数	频率（%）	频 数	频率（%）
50~60	2	5.0	2	5.0	40	100.0
60~70	7	17.5	9	22.5	38	95
70~80	12	30	21	52.5	31	77.5
80~90	14	35	35	87.5	19	47.5
90~100	5	12.5	40	100.0	5	12.5
合计	40	100.0	—	—	—	—

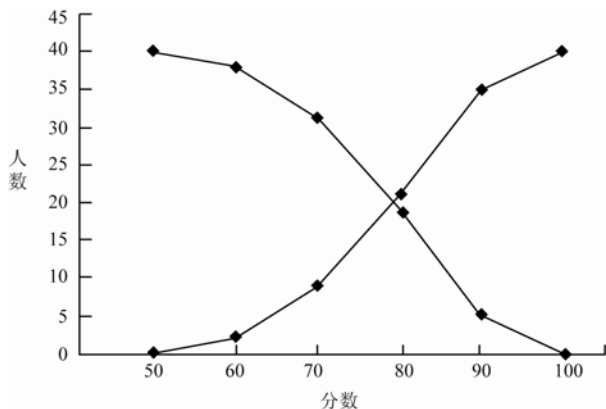


图 3.3 40 名学生英语成绩累计次数分布图

从累计次数分布表和累计次数分布曲线中，可以发现累计次数的特点：同一数值的向上累计和向下累计次数之和等于总体总次数，累计比率之和等于 100%或 1。累计次数分布还是计算位置平均数的依据。

4. 次数分布的主要类型

由于社会经济现象的复杂性，不同的统计总体有不同的次数分配，形成各种不同的次数分布类型。概括起来主要有三种类型：钟形分布、U 形分布和 J 形分布。

(1) 钟形分布。钟形分布的特征是“两头小、中间大”，即靠近中间的变量值分布的次数多，靠近两边的变量值分布的次数少，如图 3.4 所示。社会经济现象中许多变量的分布属于钟形分布，如人的身高、智商、体重，居民家庭人均收入，商品的价格等。

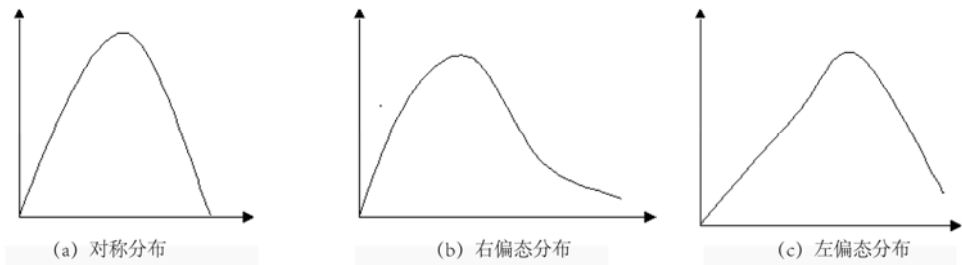


图 3.4 钟形分布

在统计学中，钟形分布根据是否对称又分为两种。一种是对称分布，也称为正态分布（normal distribution），其分布特征是中间变量值分布的次数最多，两侧变量值分布的次数随着与中间变量值距离的增大而渐次减少，两侧减少的速度相同，致使分布曲线呈对称分布，如图 3.4（a）所示。客观实际中，许多社会现象统计总体的分布都趋于对称分布中的正态分布。正态分布是描述统计中的一种主要分布，它在社会经济统计分析中具有重要的意义。另一种是非对称分布，也称为偏态分布，其分布特征是中间变量值分布的次数最多，两侧变量值的分布次数逐渐减少，但是减少的速度快慢不同，致使分布曲线向一方偏斜，呈偏斜分布（skewed distribution）。根据分布曲线的偏斜方向将偏态分布分为两种：右偏态分布和左偏态分布，如图 3.4（b）、（c）所示。

(2) U形分布。U形分布的特征是“两头大、中间小”，靠近中间的变量值分布次数少，靠近两端的变量值分布次数多，如图3.5所示。例如，人口死亡率分布就是U形分布，幼儿和老人死亡率较高，而中、青年死亡率较低；动物的死亡率也是U形分布。

(3) J形分布。J形分布的特征是“一头大、一头小”，大部分变量值集中在某一端点分布，如图3.6所示。J形分布有两种类型：正J形分布和反J形分布。正J形分布是次数随着变量的增大而增多的，如经济学中供给曲线和投资按利润率大小分布；反J形分布是次数随着变量增大而减少的，如需求曲线。

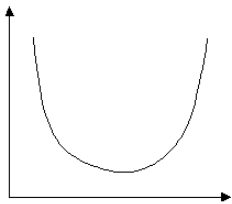


图 3.5 U形分布

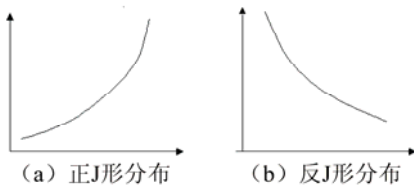


图 3.6 J形分布

3.4 统计图表

统计表和统计图是显示统计数据的重要工具。统计调查所获得的原始资料，经过统计整理，转化为能系统化的、科学的统计资料，这些统计资料往往通过统计表和统计图表示出来。

3.4.1 统计表

统计表是统计工作中应用极其广泛的一种显示统计数据的工具。

1. 统计表的定义和作用

统计表 (statistical table) 是用来显示经过汇总加工后的综合统计资料的一种表格形式。在统计工作过程中，统计调查得到的原始资料，经过统计整理后得到的可以说明社会经济现象的特征和规律的统计数据，将这些统计数据按一定的项目和顺序排列在表格上显示出来，这些表格就是统计表，如本章所编制的变量数列统计表。

采用统计表反映统计资料主要有以下优点：

- (1) 能使统计资料条理化、系统化，清晰地表达统计数据之间的相互联系。
- (2) 能简洁、明了、紧凑地显示统计数据资料，具有叙述方式显示统计数据所无法比拟的优势。
- (3) 便于计算和检查统计数据中数字的完整性和正确性。

2. 统计表的结构

统计表的结构可以从内容和形式两方面来看。

(1) 从内容上来看, 统计表包括主词和宾词两个部分。主词是就统计表所要说明的总体及其组成部分, 宾词就是统计表用来说明总体数量特征的各项统计指标及其数值。一般主词列在表的左方, 列在横栏; 宾词列在统计表的右方, 列于纵栏。但是, 统计表的主词和宾词排列的位置并不是固定不变的, 有时考虑到资料显示的某些因素, 也可以将主词与宾词的位置作以改变。

此外, 统计表还有补充资料、注解、资料来源、填表单位、填表人等附加内容。

(2) 从形式上看, 统计表包括总标题、横行标题、纵栏标题和数字资料四个部分。总标题是统计表的名称, 它简明扼要地概括了全表的基本内容, 并指明时间和范围, 一般放置于统计表格的上端正中。横行标题 (也称横标目) 是横行的名称, 就是指总体名称及其分组, 一般放在表格的左方。纵栏标题 (也称纵标目) 是纵行的名称, 就是指用于说明总体及其分组的统计指标的名称, 一般放在表格的上方。横行标题和纵栏标题共同说明填入表格中的统计数字所指的内容。数字资料是填写在统计表格的核心部分, 它就是表格中的指标数值, 用来说明总体及其组成部分的数量特征, 列在横行和纵栏的交叉处, 如图 3.7 所示。

表××				我国第六次人口普查人口性别构成表	
总标题					
横行标题	按性别分组	人数(人)	比重(%)	纵栏标题	
	男	686 852 572	51.27	数字资料	
	女	652 872 280	48.73		
	合计	1 339 724 852	100.00		
主词		宾词			

图 3.7 统计表结构

(资料来源: 国家统计局。总人口数为以 2010 年 11 月 1 日零时为标准时点普查登记的大陆 31 个省、自治区、直辖市和现役军人的人口)

由图 3.7 可知, 我国 2010 年 11 月 1 日零时大陆 31 个省、自治区、直辖市和现役军人的人口是总体, 横行标题中性别“男”和“女”是对总体的分组 (即主词), 纵栏标题中“人数”和“比重”是反映总体规模和说明总体及各组数量特征的统计指标 (即宾词), 表中的数字资料是各项指标的经济内容的具体表现。

3. 统计表的分类

根据不同的分类标志, 可以将统计表划分为不同的类型。

(1) 按用途不同, 统计表可以分为调查表、整理表和分析表。调查表是指在统计调查中用于登记、搜集和表现原始统计资料的表格。整理表是指在统计整理过程中用于统计汇总和表现汇总结果的表格。分析表是指在统计分析中用于对汇总结果进行定量分析的表格。

(2) 按总体分组的情况不同, 统计表可以分为简单表、分组表和复合表。简单表是指主词未经任何分组形成的统计表, 也称为一览表, 通常仅列出总体各单位的名称或按时间先后顺序简单排列的统计表。简单表多用于统计整理的初级阶段, 收集调查的原始资料, 如表 3.13 所示。

表 3.13 我国粮食产量情况表

年 份	粮食总产量（万吨）
2010	54 648
2011	57 121
2012	58 958
2013	60 194
2014	60 710

分组表是指主词只按某一个标志进行分组形成的统计表。利用分组表可以揭示现象不同类型的特征，反映现象的内部结构和分析现象之间的相互关系。例如，我国 2014 年年末人口年龄构成情况，如表 3.14 所示。

表 3.14 我国 2014 年年末人口年龄构成情况

按年龄划分	人数（人）	所占比重（%）
0~15 岁（含不满 16 周岁）	23 957	17.5
16~59 岁（含不满 60 周岁）	91 583	67.0
60 周岁及以上	21 242	15.5
合 计	136 782	100.0

复合表是指主词按两个或两个以上的标志进行重叠式分组形成的统计表。利用复合表可以反映所研究的现象受几种因素的共同影响而发生的变化，可以更加准确地把握现象变化的规律，详细地认识问题和说明问题，如表 3.15 所示。然而，复合分组的标志并不是越多越好，因为分组标志越多，组数就会成倍增加，而且分组太细也不利于研究现象的特征。

表 3.15 北京市 2014 年年末常住人口及构成表

指标	人数（万人）	比重（%）
常住人口	2 151.6	100.0
按城乡分：城镇	1 859	86.4
乡村	292.6	13.6
按性别分：男性	1 106.5	51.4
女性	1 045.1	48.6
按年龄组分：0~14 岁	213.0	9.9
15~59 岁	1 617.0	75.2
60 岁及以上	321.6	14.9
其中：65 岁及以上	212.3	9.9
按功能区分：首都功能核心区	221.3	10.3
城市功能拓展区	1 055.0	49.0
城市发展新区	684.9	31.8
生态涵养发展区	190.4	8.9

（资料来源：2014 年国民经济和社会发展统计公报）

4. 统计表的设计

统计表设计一般是按宾词进行设计。统计表的设计分为简单设计和复合设计。简单设计就是将所要反映的指标平行排列，如表 3.16 所示。简单设计通过平行分组直接说明指标的内容。

复合设计是将各个指标结合起来进行层叠排列，如表 3.17 所示。

表 3.16 某高校某年招收的新生性别与城乡状况统计表（1）

按学历分组	学 生 人 数	性 别		城 乡	
		男	女	城 镇	农 村
甲	(1)	(2)	(3)	(4)	(5)
专科					
本科					
合 计					

表 3.17 某高校某年招收的新生性别与城乡状况统计表（2）

按学历分组	学 生 人 数			城 乡					
				城 镇			农 村		
	男	女	小计	男	女	小计	男	女	小计
甲	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
专科									
本科									
合 计									

5. 统计表编制的规则

为了使统计表能更好地反映被研究对象的数量特征，便于比较分析，应遵循以下编制规则，使统计表更加科学、实用、简明、美观。

1) 统计表表式设计规则

① 表格样式。统计表应设计成由纵横交叉线条组成的长方形表格，长与宽之间保持适当的比例。统计表的上下两端的端线应当用粗线绘制，表中其他线条一律用细线绘制，表的左右两端习惯上均不画线，采用开口式。

② 合计栏的设置。表中的横行“合计”，一般列在最后一行（或最前一行）；表中纵栏的“合计”一般列在最前一栏。

③ 栏数的编号。如果栏数较多，应当按顺序编号，习惯上主词栏部分分别编以“甲、乙、丙、丁……”序号，宾词栏编以“(1)、(2)、(3)……”序号。

2) 统计表内容设计规则

① 标题设计。无论是总标题还是横行、纵栏标题，都应简明扼要，简练而又准确地表述出统计资料的内容及所属的时间和空间范围。



② 标题顺序。统计表中的主词与宾词的排列应尽量反映出内容方面的逻辑关系。

③ 指标数值。表中数字应填写整齐，上下位置对齐。遇有相同数字应照写，不能用“同上”、“同左”等字样；当数字因小而忽略不计时，可填写为“0”；当缺某项数字资料时，可用符号“...”表示；不应有数字时，用符号“-”表示。

④ 计量单位。指标数字应有计算单位。如果全表的计算单位是相同的，可以把它写在表头的右上方；如果表中指标数字计算单位各栏之间不相同，应在各栏标题中注明计算单位。

⑤ 注解与资料来源。为保证统计资料的科学性与严肃性，在统计表下，应注明资料来源，以便考察。必要时，在统计表下加注说明。

3.4.2 统计图

1. 统计图的作用和结构

统计表的资料用几何图形或图案等形式表示即成为统计图。统计图也是显示统计数据的一种形式。如果说利用统计表能够集中有序地表现统计资料，那么利用统计图则能够更加形象具体、简明生动、一目了然地展示统计资料，便于人们直观地认识事物的特征。统计图可以表示现象之间的对比关系，揭示总体结构及其变化发展趋势，分析现象之间的依存关系等。随着计算机技术不断发展，电脑制图功能日益强大，使得统计图的制作更加方便和精确。

统计图由标题、图域、标目、尺度和图例五部分构成。标题，每个图都应有标题，标题要简明确切，通常包括内容、时间和地点，其位置在图域之外，一般放在图域的下面。图域，图域的长宽之比一般 7:5 为美观，圆图除外。标目，纵、横两轴应有标目，即纵标目和横标目，并注明度量衡单位。尺度，纵、横两轴都有尺度，横轴尺度自左至右，纵轴尺度自下而上，数值一律由小到大，尺度间隔要宽松，用算术尺度时，等长的距离应代表相等的数量。图例，用不同线条或颜色代表不同事物时，需用图例说明。

2. 常见的统计图的绘制

统计图（statistical graph）的形式多种多样，本节主要介绍以下常用的三种图形。

1) 条形图

条形图（bar chart）也称为柱形图，是用宽度相同的条形的高度或长度来表示统计数据大小或多少的一种图形。条形图可以横置也可以纵置，纵置时又称为柱形图。也就是说，当各类别放在纵轴时，称为条形图；当各类别放在横轴时，称为柱形图。它主要用于说明或比较同一指标在不同时间、地点、单位的变化发展情况。例如，表 3.18 与图 3.8 反映了我国 2004—2014 年国内游客人数的变动情况。

表 3.18 我国 2004—2014 年国内游客人数表

年 份	国内游客（百万人次）
2004	1 102
2005	1 212
2006	1 394
2007	1 610
2008	1 712
2009	1 902
2010	2 103
2011	2 641
2012	2 957
2013	3 262
2014	3 610

（资料来源：中国统计年鉴 2015）

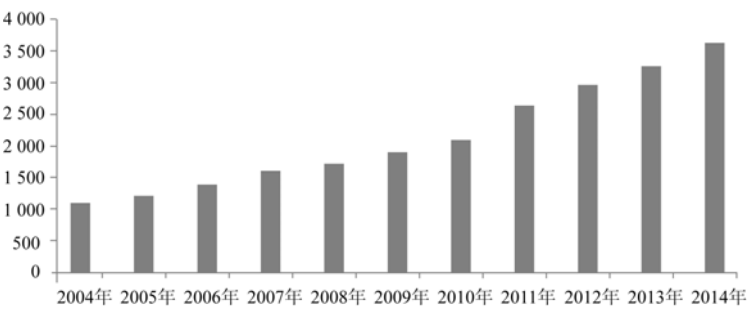


图 3.8 我国 2004—2014 年国内游客人数条形图

2) 圆形图

圆形图（circle chart）又称为饼图（pie graph），是用圆形和圆内扇形的面积大小来表示统计指标数值大小的一种图形。它用于表示总体中各组成部分所占的比例，揭示现象的内部结构及其变化。在绘制圆形图时，总体中各部分所占的百分比用圆内的各个扇形面积表示，这些扇形的中心角度是按各部分百分比占 360° 的相应比例确定的。如图 3.9 所示，显示了我 国 2014 年年末人口按年龄分布的状况。

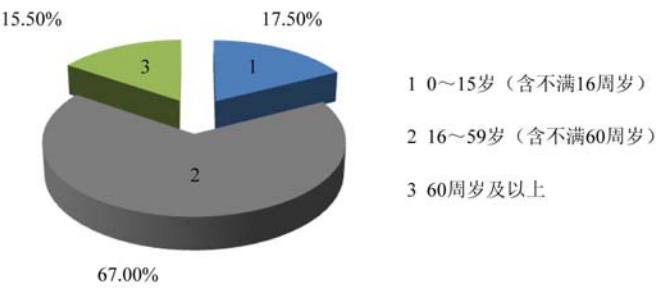


图 3.9 2014 年年末我国人口年龄构成饼图

3) 曲线图

曲线图（curvilinear chart）又称为折线图，是利用曲线的升、降变化来表示统计指标数值变化的一种图形。它用于分析社会经济现象的发展变化规律、趋势及现象之间的依存关系。

绘制曲线图时，如果是某一现象随时间变化的显示，则应将时间绘制在横坐标轴上，指标绘制在纵坐标轴上；如果是两个现象依存关系的显示，则一般将表示原因的指标绘制在横坐标轴上，表示结果的指标绘制在纵坐标轴上。

例如，2005—2014 年我国城乡居民人均收入，如表 3.19 和图 3.10 所示。

表 3.19 2005—2014 年我国城乡居民人均收入

年 份	城镇居民人均可支配收入（元）	农村居民人均纯收入（元）
2005	10 493.0	3 254.9
2006	11 759.5	3 587.0
2007	13 785.8	4 140.4
2008	15 780.8	4 760.6
2009	17 174.7	5 153.2
2010	19 109.4	5 919.0
2011	21 809.8	6 977.3
2012	24 564.7	7 916.6
2013	26 955.1	8 895.9
2014	28 844.0	9 892.0

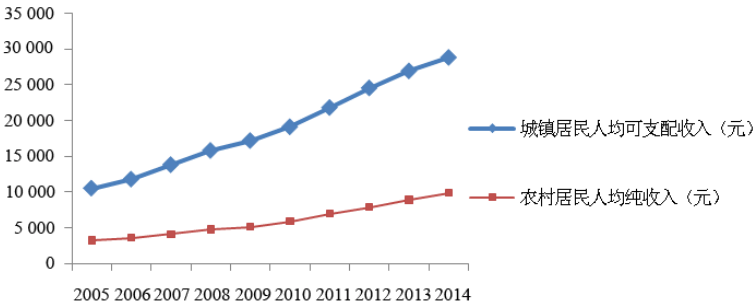


图 3.10 2005—2014 年我国城乡居民人均收入曲线图

从图 3.10 中可以清楚地看出，城乡居民的人均收入逐年提高，城镇居民人均可支配收入高于农村居民人均纯收入，而且这种差距有扩大的趋势。

绘制曲线图时应注意：图形的长宽比例要适当，一般为横轴略大于纵轴的长方形，其长宽比例大致为 10：7，图形过扁或过于瘦高，不仅不美观，而且还会给人造成视觉上的错觉，不利于对数据变化的理解。

3. 绘制统计图的基本要求和原则

为了使统计图能准确生动地反映被研究对象的数量特征，在编制统计图时应注意以下编制规则：

- (1) 各种图形的适用条件不同, 因此, 要选择恰当的图形来表示统计数据。
- (2) 图的标题要简明扼要地说明所要表达的内容。图的标题位于表的下方, 字体是最大的。
- (3) 有纵横轴的图形, 横轴表示研究对象, 尺度要等距, 自左至右, 由小到大; 纵轴一般表示现象出现的频数或频率, 从零开始等距分点, 由下至上, 从小到大。数字位于左侧, 要注明单位。
- (4) 线条粗细有差别, 图形线条最粗, 坐标线条较细。
- (5) 在同一图形上比较多个事物时取的尺度要相同, 比较的对象不宜过多。
- (6) 图形上尽量不要写数字, 如要说明则应该归于图注中。

3.5 Excel在数据整理中的应用

通过统计调查得到的数据是杂乱的, 没有规则的, 因此, 必须对搜集到的大量的原始数据加工整理, 经过数据分析得到科学结论。统计整理包括对数据进行分类汇总并计算各类指标及利用统计图或统计表描述统计汇总结果等。Excel 提供了多种数据整理工具: 频数分布函数、统计图、直方图分析工具等。本节操作以【实例 3.4】为操作范例进行阐述。

【实例 3.4】 根据抽样调查, 某月某市 50 户居民购买消费品支出资料如下 (单位: 元), 对其按 800~900、900~1 000、1 000~1 100、1 100~1 200、1 200~1 300、1 300~1 400、1 400~1 500、1 500~1 600、1 600 以上分为 9 个组。

830	880	1 230	1 100	1 180	1 580	1 210	1 460	1 170	1 080
1 050	1 100	1 070	1 370	1 200	1 630	1 250	1 360	1 270	1 420
1 180	1 030	870	1 150	1 410	1 170	1 230	1 260	1 380	1 510
1 010	860	810	1 130	1 140	1 190	1 260	1 350	930	1 420
1 080	1 010	1 050	1 250	1 160	1 320	1 380	1 310	1 270	1 250

3.5.1 利用Excel进行统计分组

Excel 提供了一个专门用于统计分组的频数分布函数 (FREQUENCY), 它以一系列垂直数组返回某个区域中的数据分布, 描述数据分布状态。用频数分布函数进行统计分组的操作过程如下。

首先, 在使用此函数时, 先将样本数据排成一列, 本例中为 A1:A50。

然后, 利用频数分布函数进行统计分组和计算频数, 具体操作步骤如下。

第一步: 在选定单元格区域, 单击“插入”菜单, 选择“函数”选项, 弹出“粘贴函数”对话框。在对话框的左侧“函数分类”中选择“统计”, 在右侧的“函数名”中选择 FREQUENCY, 如图 3.11 所示。本例中选定的单元格区域为 C3:C11。

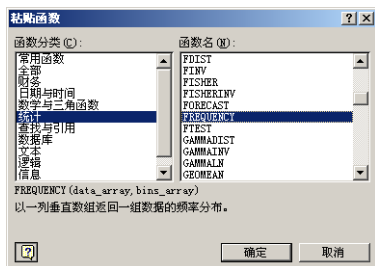


图 3.11 “粘贴函数”对话框

第二步：打开 FREQUENCY 对话框，输入待分组数据与分组标志，如图 3.12 所示。在 FREQUENCY 对话框中有“Data_array”和“Bins_array”两个框。

(1) Data_array 用于计算频率的数组，或对数组单元区域的引用。本例中为 A1:A50。

(2) Bins_array 数据接收区间，为一组数或对数组区间的引用，设定对 Data_array 进行频率计算的分段点。本例中为 899、999、1 099、1 199、1 299、1 399、1 499、1 599、1 699。

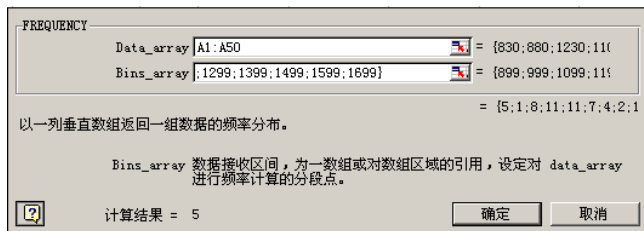


图 3.12 FREQUENCY 对话框

注意：频数分布函数要求按组距的上限分组，不接受非数值的分组标志（如“不足××”或“××以上”等）。在输入的数据两端必须加大括号，各数据之间用分号隔开。输入完成后，由于频数分布是数组操作，所以不能按“确定”按钮。

第三步：按“Ctrl+Shift+Enter”组合键，在最初选定单元格区域内得到频数分布结果，在本例中为 C3:C11，如图 3.13 所示。

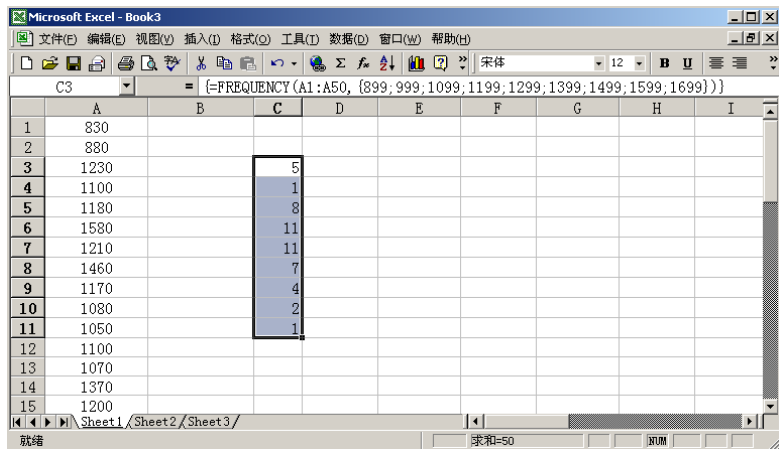


图 3.13 频数分布结果

至此，频数分布函数进行统计分析的功能就全部完成了。如需绘制频数分布表，则需配合 Excel 的其他功能完成。

在利用频数分布函数统计分组获得频数的基础上，再利用 Excel 列表计算频数分布表，如图 3.14 所示。

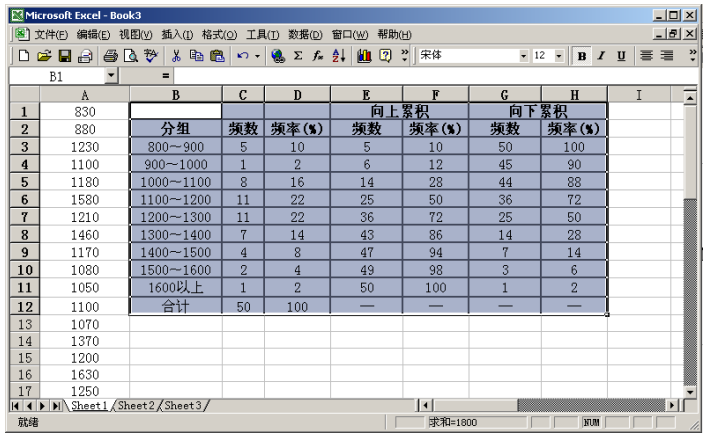


图 3.14 频数分布表

利用函数和公式可获得频数分布表。计算方法如下：C 列频数之和可利用 SUM 求和函数，单击 C11 单元格，输入“=SUM（C3:C11）”，回车得 C11 结果。D 列频数，单击 D3 单元格，输入“=C3*100/50”，回车得 D3 结果，然后，使用填充柄功能按住鼠标左键向下拖，至 D11 单元格，放开鼠标，可得 D4～D11 结果；求和类似于频数之和的做法。E 列向上累积频数，单击 E3 单元格，输入“=C3”，回车可得结果，然后再单击 E4 单元格，输入“=E3+B4”，回车得结果，然后，使用填充柄功能按住鼠标左键向下拖，至 E11 单元格，放开鼠标，可得 E4～E11 结果；F 列向上累积频率的计算类似于向上累积频数计算。G 列向下累积频数，单击 G3 单元格，输入“=C12”，回车可得结果，然后再单击 G4 单元格，输入“=G3-C3”，回车得结果，然后，使用填充柄功能按住鼠标左键向下拖，至 G11 单元格，放开鼠标，可得 G4～G11 结果；H 列向上累积频率的计算类似于向上累积频数计算。

与频数分布函数只能进行统计分组和频数计算相比，直方图分析工具可完成数据的分组、频数分布与累积频数的计算、绘制直方图与累积折线图等一系列操作。仍以例 3.4 阐述直方图分析工具的统计整理功能，其操作过程如下。

首先，先将样本数据排成一列，最好对数据进行排序，本例中已利用排序操作排好序，为 A1:A50。输入分组标志，本例中为 B1:B10，分别是 899、999、1 099、1 199、1 299、1 399、1 499、1 599、1 699。

注意：要求按组距的上限分组，不接受非数值的分组标志（如“不足××”或“××以上”等）。

然后，利用直方图分析工具进行分析，具体操作步骤如下。

第一步：单击“工具”菜单，选择“数据分析”选项，在弹出的“数据分析”对话框中，选择“直方图”，如图 3.15 所示。

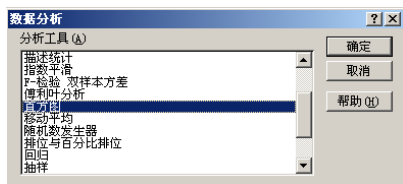


图 3.15 “数据分析”对话框

第二步：打开“直方图”对话框，确定输入区域、接收区域和输出区域，如图 3.16 所示，在“直方图”对话框中有“输入”和“输出选项”。

(1) “输入区域”输入待分析数据区域的单元格引用，若输入区域有标志项，则勾选“标志”复选框；否则，系统自动生成数据标志。“接收区域”输入接收区域的单元格引用，该框可为空，则系统自动利用输入区域中的最小值和最大值建立平均分布的区间间隔的分组。本例中，输入区域为\$A\$1:\$A\$51，接收区域为\$B\$1:\$B\$10。

(2) 在“输出选项”中可选择输出去向，输出去向类似于“抽样”对话框的输出去向。本例中选择“输出区域”，为\$C\$1。

(3) 选择“柏拉图”可以在输出表中同时按降序排列频数数据；选择“累积百分率”可在输出表中增加一列累积百分比数值，并绘制一条百分比曲线；选择“图表输出”可生成一个嵌入式直方图。

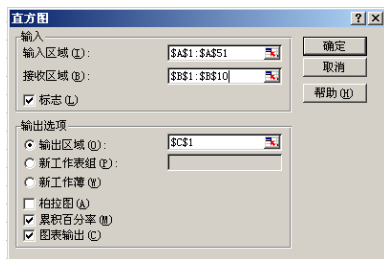


图 3.16 “直方图”对话框

第三步：单击“确定”按钮，在输出区域单元格得到频数分布，如图 3.17 所示。

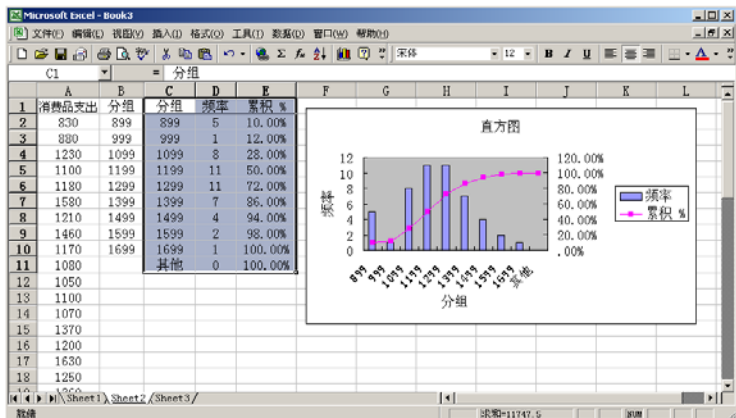


图 3.17 直方图频数分布结果

第四步：将条形图转换成标准直方图，如图 3.18 所示。具体做法：单击条形图的任一直条，再单击鼠标右键，在快捷菜单中选择“数据系列格式”，然后在弹出的“数据系列格式”对话框中选择“选项”标签，将间距宽度改为 0，单击“确定”按钮即可。

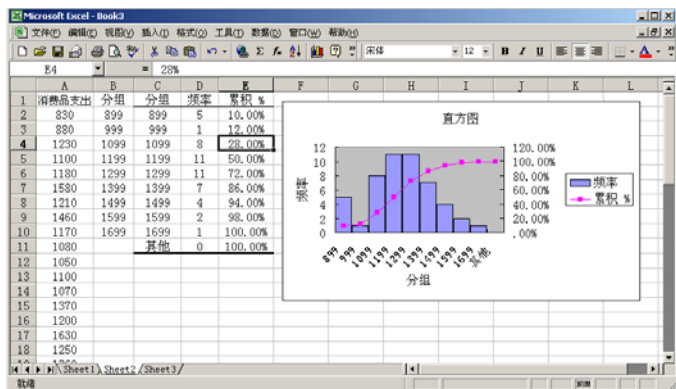


图 3.18 标准直方图

3.5.2 利用Excel绘制统计图

传统的统计表格需要数据使用者精心分析，而统计图显示资料具有形象生动、一目了然的优点，通过图形可以方便地观察到数量之间的对比关系、总体的结构特征及变化发展趋势等。统计图在统计整理中的应用越来越广泛。Excel 提供了大量的统计图形供用户根据需要进行选择和图形功能选择使用。Excel 提供的图形工具有：柱形图、折线图、饼图、散点图、面积图、环形图、股价图等。利用 Excel 制作统计图的工作步骤如下，以柱形图为例。

【实例 3.5】 根据我国 1991—2005 年国内生产总值构成数据制作相应的统计图。

第一步：创建工作表，即将统计资料输入到 Excel 中，如图 3.19 所示。

年份	第一产业	第二产业	第三产业
1991	24.3	41.8	33.9
1992	21.5	43.5	35.0
1993	19.5	46.6	33.9
1994	19.6	46.6	33.8
1995	19.8	47.2	33.0
1996	19.5	47.5	33.0
1997	18.1	47.5	34.4
1998	17.3	46.2	36.5
1999	16.2	45.8	38.0
2000	14.8	45.9	39.3
2001	14.1	45.2	40.7
2002	13.5	44.8	41.7
2003	12.6	46.0	41.4
2004	13.1	46.2	40.7
2005	12.6	47.5	39.9

图 3.19 1991—2005 年国内生产总值构成数据

第二步：在想绘制图形的数据中选择任一单元格，然后在“插入”菜单中选择“图表”选项，或单击工具栏中的图表向导按钮，如图 3.20 所示。



图 3.20 选择“图表”选项

第三步：选定图表类型。在弹出的“图表向导-4 步骤之 1-图表类型”对话框中选择所需要的图表类型，然后单击“下一步”按钮，如图 3.21 所示。



图 3.21 “图表向导-4 步骤之 1-图表类型”对话框

第四步：确定数据范围。在弹出的“图表向导-4 步骤之 2-图表源数据”对话框中规定数据区域和每一数据系列的名字和数值的区域，如图 3.22 和图 3.23 所示，然后单击“下一步”按钮。



图 3.22 “图表向导-4 步骤之 2-图表源数据”对话框

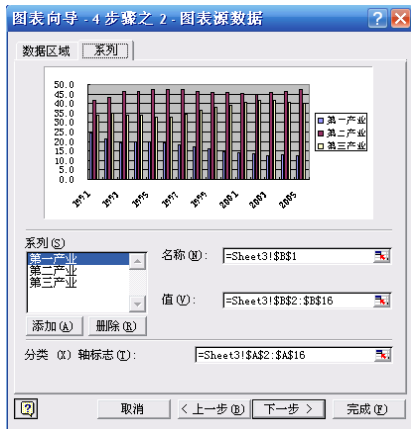


图 3.23 “图表向导-4 步骤之 2-图表源数据”数据区域选项

第五步：选用图表选项。在弹出的“图表向导-4 步骤之 3-图表选项”对话框中可给图表命名、为每一个坐标轴赋予名称等，然后单击“下一步”按钮，如图 3.24 所示。

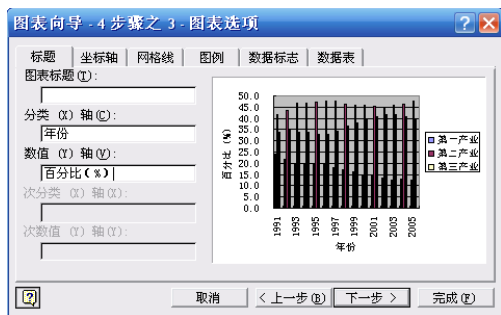


图 3.24 “图表向导-4 步骤之 3-图表选项”对话框

第六步：选择图表位置并显示结果。在弹出的“图表向导-4 步骤之 4-图表位置”对话框中可为图表选择保存位置，或放在独立的工作表中，或作为一个对象放在当前工作表中，然后单击“完成”按钮。本例中选择了作为一个对象放在工作表中，如图 3.25 所示。

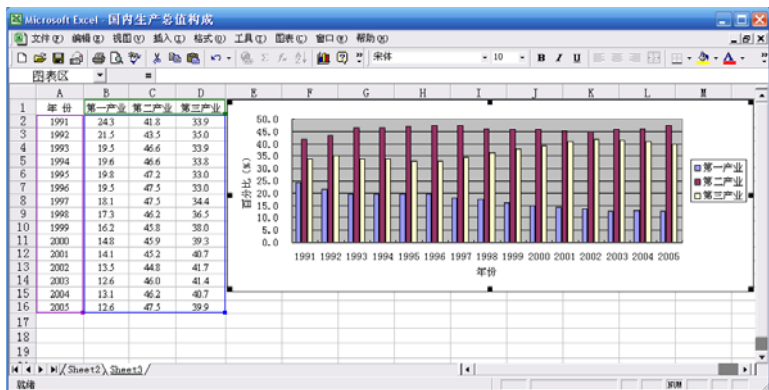


图 3.25 1991—2005 年国内生产总值构成图表

其他图形的绘制步骤与柱形图的绘制步骤类似，只要在图形类型中选择不同的图形类型即可。

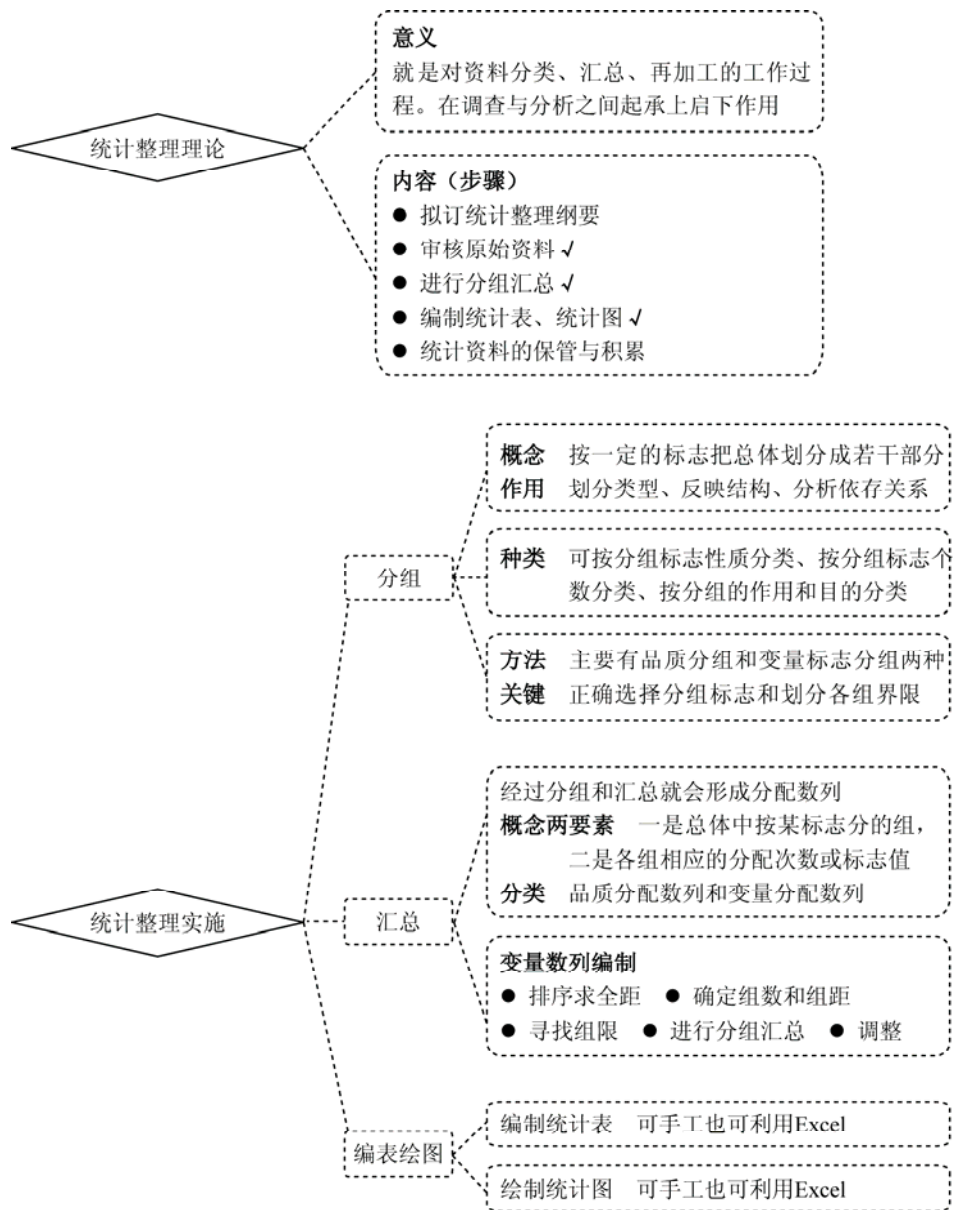


统计术语

统计整理 statistical treatment	统计分组 statistical classification
频数分布 frequency distribution	频数 frequency
组距 class interval	上限 upper limit
下限 lower limit	组中值 mid-value of class
直方图 histogram	正态分布 normal distribution
偏斜分布 skewed distribution	统计表 statistical table
统计图 statistical graph	条形图 bar chart
曲线图 curvilinear chart	



重点知识梳理



习题与实践训练

一、判断题

1. 统计分组的关键是确定组距和组数。 ()
2. 离散型变量只适合于单项式分组。 ()
3. 凡是将总体按某个标志值分组所形成的数列，都叫变量数列。 ()

4. 按数量标志分组的目的是要区分各组在数量上的差别。 ()
5. 连续性变量只能编制组距式变量数列; 离散型变量既可以编制单项式变量数列, 也可以编制组距式变量数列。 ()
6. U 型分布的特征是“两头大、中间小”。 ()
7. 变量数列中的开口组不能确定组中值。 ()
8. 累计次数的特点: 同一数值的向上累计和向下累计次数之和等于 1。 ()
9. 统计分组以后, 掩盖了各组内部各单位的差异, 而突出了各组之间单位的差异。 ()
10. 连续型变量和离散型变量在进行组距式分组时, 均可采用相邻组组距重叠的方法确定组限。 ()

二、单项选择题

1. 某连续变量数列, 其首组为 100 以下, 若其相邻的组中值为 130, 则首组的组中值为 ()。
- A. 60 B. 70 C. 80 D. 100
2. 变量数列中各组频率之和应该是 ()。
- A. 不等于 1 B. 等于 1 C. 小于 1 D. 大于 1
3. 统计整理中, 最关键的工作是 ()。
- A. 审核原始资料 B. 资料汇总 C. 统计分组 D. 编制统计表和统计图
4. 统计分组的目的是为了体现 ()。
- A. 组内同质性、组间差异性 B. 组内同质性、组间同质性
- C. 组内差异性、组间同质性 D. 组内差异性、组间差异性
5. 连续变量分组, 若第一组为 100 以下, 第二组为 100~150, 第三组为 150~200, 第四组为 200 以上, 则数据 ()。
- A. 100 在第一组 B. 150 在第二组 C. 200 在第三组 D. 200 在第四组
6. 次数密度是 ()。
- A. 各组单位组距内分布次数 B. 平均每组组内分布的频率
- C. 单位次数的组距长度 D. 平均每组组内分布的次数
7. 下列属于数量标志分组的是 ()。
- A. 某工业企业工人按民族分组
- B. 某工业企业工人按性别分组
- C. 某工业企业工人按技术职称分组
- D. 工业企业按年龄分组
8. 企业按资产总额分组 ()。
- A. 只能使用单项式分组
- B. 只能使用组距式分组
- C. 可以单项式分组, 也可以用组距式分组
- D. 无法分组
9. 次数分配数列是 ()。
- A. 按数量标志分组形成的数列

- B. 按品质标志分组形成的数列
 - C. 按统计指标分组所形成的数列
 - D. 按数量标志和品质标志分组所形成的数列
10. 在组距分组时, 对于连续型变量, 相邻两组的组限 ()。
- A. 必须是重叠的
 - B. 必须是间断的
 - C. 可以是重叠的, 也可以是间断的
 - D. 必须取整数

三、多项选择题

1. 指出下列分组哪些是按品质标志分组 ()。
 - A. 企业按所有制分组
 - B. 家庭按人口多少分组
 - C. 人口按居住地区分组
 - D. 固定资产按用途分组
 - E. 职工按工资水平分组
2. 统计分组的主要作用有 ()。
 - A. 说明总体单位的数量特征
 - B. 反映总体内部结构
 - C. 研究现象之间的依存关系
 - D. 划分现象的类型
 - E. 反映总体的基本情况
3. 在组距数列中, 组中值是 ()。
 - A. 上限和下限之间的中点数值
 - B. 用来代表各组标志值的平均水平
 - C. 在开放式分组中无法确定
 - D. 就是组平均数
 - E. 在开放式分组中, 可以参照相邻组的组距来确定
4. 在次数分配数列中 ()。
 - A. 总次数一定, 频数和频率成反比
 - B. 各组的频数之和等于 100
 - C. 各组频率大于 0, 频率之和等于 1
 - D. 频率又称为次数
 - E. 频数越小, 则该组的标志值所起的作用越小
5. 统计表按分组的情况可分为 ()。
 - A. 简单表
 - B. 调查表
 - C. 分组表
 - D. 整理表
 - E. 复合表
6. 正确选择分组标志的原则是 ()。
 - A. 要根据事物发展的规律选择分组标志
 - B. 选择最能体现事物本质特征的标志作为分组标志
 - C. 要根据研究的目的和任务选择分组标志
 - D. 根据数量标志和品质标志不同来选择分组标志
 - E. 要结合现象的历史条件和经济条件来选择分组标志
7. 审核资料的主要内容包括对 () 的审核。
 - A. 资料的广泛性
 - B. 资料的准确性
 - C. 资料的及时性
 - D. 资料的完整性
 - E. 资料的规范性
8. 从内容来看, 所有的统计表都是由哪几个部分组成 ()。
 - A. 横行标题
 - B. 纵栏标题
 - C. 宾词
 - D. 主词
 - E. 数字资料
9. 下列只能编制组距数列的有 ()。

- A. 家庭按拥有微机数分组

B. 职工按月工资额分组
- C. 商场按营业收入分组

D. 学生按每周上网时间分组
- E. 城市按年地区生产总值分组
10. 某一分配数列如下:

实习员工每天拼装玩具数（个）	员工人数（人）
5 以下	7
5~9	18
10~14	52
15~19	20
20 以上	3
合 计	100

- 以下说法正确的是（ ）。
- A. 第四组的组中值是 17

B. 最后一组为开口组
- C. 第三组的组距是 5

D. 第一组下限是 5
- E. 第五组的频率是 30%

四、填空题

1. 统计分组的关键是_____。
2. 按分组的作用或目的不同，统计分组可划分为_____、_____和_____。
3. 分配数列按分组标志特征的不同，分为_____和_____。
4. 在组距数列中，表示各组界限的变量值称为_____，下限是指_____的变量值，上限是指_____的变量值。各组上限与下限之间的中点数值称为_____。
5. 组距与组数关系密切，同一现象的变量分组，组距与组数成_____。
6. 连续变量分组，相邻两组的上限与下限通常是相同的，为了避免计算时总体单位数出现错误，统计各组单位数时遵循这一原则：将到达上限值的单位数计入下一组内，这一原则被称为_____。
7. 组距式分组根据其分组的组距是否相等，可分为_____和_____。
8. 次数分配是由_____和_____两个要素构成的。表示各组单位数的次数又称为_____，各组次数与总次数之比称为_____。
9. 各种不同性质的社会经济现象的次数分布的类型，根据曲线形状的特点主要有三种类型：_____、_____、_____。
10. 统计表从表式上看，包括_____、_____、_____和_____四个部分。

五、应用能力训练题

1. 某集团下属 32 家公司青年员工人数资料如下：

322	674	339	357	346	295	465	355
332	316	453	442	417	587	369	545
323	430	560	528	333	311	410	604
281	461	432	421	484	473	392	354

根据上面的数据编制组距为 100 的分布数列，并绘制次数分布图。

2. 为了确定灯泡的使用寿命（小时），在一批灯泡中随机抽取 100 只进行测试，得到结果如下。

700	716	728	719	685	709	691	684	705	718
706	715	712	722	691	708	690	692	707	701
708	729	694	681	695	685	706	661	735	665
668	710	693	697	674	658	698	666	696	698
706	692	691	747	699	682	698	700	710	722
694	690	736	689	696	651	673	749	708	727
688	689	683	685	702	741	698	713	676	702
701	671	718	707	683	717	733	712	683	692
693	697	664	681	721	720	677	679	695	691
713	699	725	726	704	729	703	696	717	688

根据上述资料：

- (1) 整理成组距数列（以组距为 10 进行等距分组），编制分布表；
- (2) 绘制直方图和次数分配曲线。

3. 某班学生统计学考试成绩（分）如下：

93	50	78	85	66	71	63	83	52	95
78	72	85	78	82	90	80	55	95	67
72	85	77	70	90	70	76	69	58	89
80	61	67	99	89	63	78	74	82	88
98	62	81	44	76	86	73	83	85	81

根据上述资料：

- (1) 编制组距数列，说明每一组的上下限、组中值。
- (2) 绘制次数分布的曲线图，据此分析成绩分布的特点。
- (3) 编制累计频数分布表，并回答 60 分以下及 80 分以上的人数。
4. 针对下面的每一种情况，说明当你要使用图表进行描述时，你是使用饼图、趋势图还是条形图等，为什么？

- (1) 人力资源部门统计申报五个不同岗位的人数；
- (2) 某学院男女学生的比重；
- (3) 2005—2015 年公司每年主营业务收入数值；
- (4) 八个项目中每个项目的得分。

5. 根据对所在大学的学生进行上网时间调查所获得的原始资料进行统计整理，设计各种统计表并绘制统计图，以反映学生上网时间、上网目的、上网途径等的分布情况，在此基础上了解学生上网的综合情况，并编写统计调查分析报告。

6. 利用第 2 章中“应用能力训练题第 5 题”所搜集的资料：

- (1) 按性别分组，编制品质分配数列。
- (2) 按年龄分组，编制单项式变量数列。

(3) 按身高分组, 编制组距式变量数列。

7. 从生产车间 500 名工人中随机抽取 50 名工人, 50 名工人的日产量如下所示 (单位: 件)。

148	140	127	120	110	104	132	135	129	123
116	109	132	135	129	123	110	108	148	135
128	123	114	108	132	124	120	125	116	118
125	137	107	113	123	140	137	119	119	127
128	119	124	130	118	107	113	122	128	114

根据上述资料:

(1) 利用 Excel 频数分布函数进行描述分布状态, 分为 5 组, 计算频率和累积频数, 绘制直方图和累积折线图。

(2) 利用 Excel 直方图工具进行数据分组, 计算频数、频率和累积频数, 给出直方图和累积折线图。

8. 去图书馆找一篇你感兴趣的领域的期刊文章, 文章包含一些数据但是没有对数据的图形表示。利用这些数据设计统计表、绘制统计图。你要建立哪一类型的图表? 为何作出这样的选择? 你可以手绘统计图表, 也可使用 Excel 建立图表。



本章案例

第 36 次中国互联网络发展状况统计报告

一、网民规模

(一) 总体网民规模

截至 2015 年 6 月, 我国网民规模达 6.68 亿, 半年共计新增网民 1 894 万人。互联网普及率为 48.8%, 较 2014 年年底提升了 0.9 个百分点, 整体网民规模增速继续放缓, 见图 1。

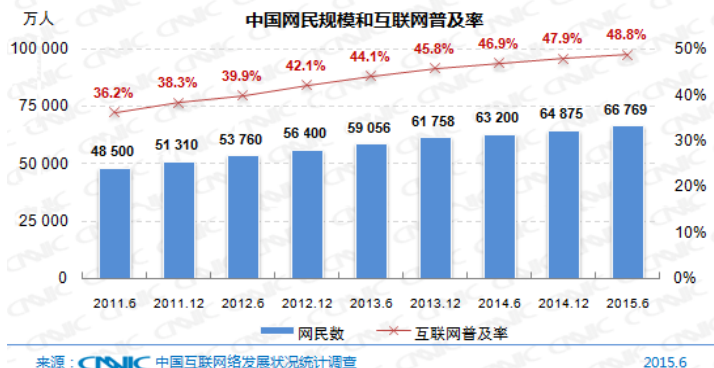


图 1 中国网民规模和互联网普及率图

随着网民规模的增长进入平台期, 互联网对个人生活方式的影响进一步深化, 从基于信息获取和沟通、娱乐需求的个性化应用, 发展到与医疗、教育、交通等公用服务深度融合的民生服务。与此同时, 随着“互联网+”行动计划的出台, 互联网将带动传统产业的变革和创新。未

来，在云计算、物联网及大数据等应用的带动下，互联网将加速农业、现代制造业和生产服务业转型升级，形成以互联网为基础设施和实现工具的经济发展新形态。

（二）手机网民规模

截至2015年6月，我国手机网民规模达5.94亿，较2014年12月增加了3679万人。网民中使用手机上网的人群占比由2014年12月的85.8%提升至88.9%，见图2。



图2 中国手机网民规模及其占网民比例

移动上网设备的逐渐普及、网络环境的日趋完善、移动互联网应用场景的日益丰富三个因素共同作用，促使手机网民规模进一步增长。

首先，智能手机价格下降，为手机上网奠定了基础。2015年上半年，各大互联网厂商和传统家电企业对于手机市场的进入，促使智能手机价格持续走低，提升了网民购买力。其次，政府加大对于移动上网的扶持，通过督促运营商降低上网资费、提升网络覆盖能力等措施优化网民上网环境，降低手机上网门槛。最后，移动互联网应用场景的丰富提升了网民使用意愿。移动互联网与传统行业加速融合，开发与各类生活紧密关联的新应用，吸引传统行业用户开始使用移动互联网。

（三）农村网民规模

截至2015年6月，我国网民中农村网民占比为27.9%（见图3），规模达1.86亿，相比2014年年底增加了800万人。

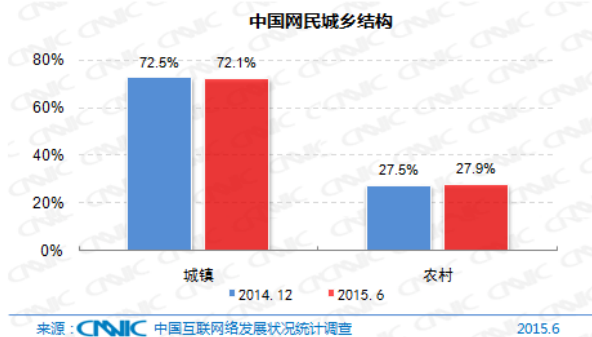


图3 中国网民城乡结构图

城镇地区与农村地区的互联网普及率分别为64.2%、30.1%，相差34.1个百分点。人口结

构方面，10~40 岁人群中，农村地区的互联网普及率比城镇地区低 15~27 个百分点（见图 4），这部分人群互联网普及的难度相对较低，将来可转化的空间较大。

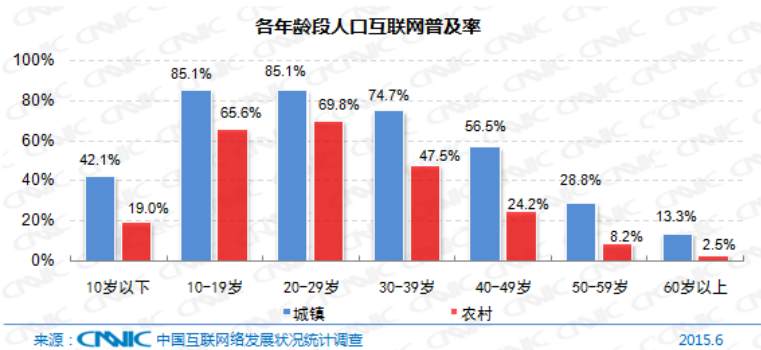


图 4 各年龄段人口互联网普及率

未来农村地区互联网发展将由政府与互联网企业共同带动。政府方面，应加大互联网基础设施建设及政策扶持，提升农村人口对互联网的认知及使用。2015 年 6 月国务院办公厅印发《关于支持农民工等人员返乡创业的意见》，鼓励输出地资源对接输入地市场带动返乡创业。在这一过程中，农民工可以发挥既熟悉输入地市场又熟悉输出地资源的优势，借力“互联网+”信息技术发展现代商业，实现输出地产品与输入地市场的对接，进而带动农村地区互联网的发展。企业方面，需要针对农村地区特性提供更贴近农村地区需求的应用，提升农村人口使用互联网的意愿。目前阿里、京东、腾讯等互联网企业纷纷推出针对农村地区的农业电商和农村金融服务，这些举措将对农村互联网发展起到带动作用。

二、网民属性

（一）性别结构

截至 2015 年 6 月，中国网民男女比例为 55.1 44.9，女性网民占比相比 2014 年年底提升了 1.3 个百分点，见图 5。

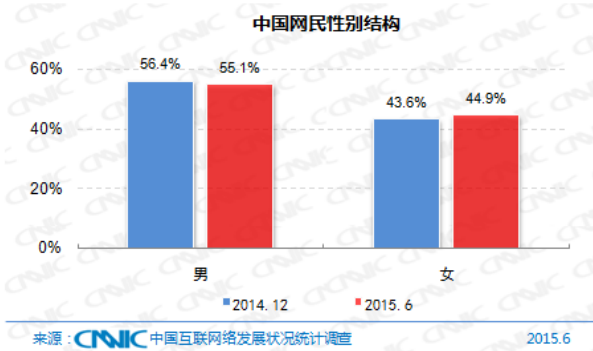


图 5 中国网民性别结构图

（二）年龄结构

截至 2015 年 6 月，我国网民以 10~39 岁年龄段为主要群体，比例达到 78.4%。其中，20~29 岁年龄段网民的比例为 31.4%，在整体网民中的占比最大（见图 6）。与 2014 年年底相比，20 岁以下网民规模占比增长 1.1 个百分点，互联网继续向低龄群体渗透。

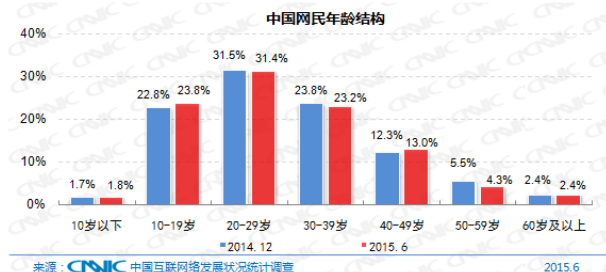


图6 中国网民年龄结构图

(三) 学历结构

截至2015年6月,整体网民中小学及以下学历人群的占比为12.4%,较2014年年底上升1.3个百分点。与此同时,大专及以上学历占比下降0.8个百分点,中国网民继续向低学历人群扩散,见图7。

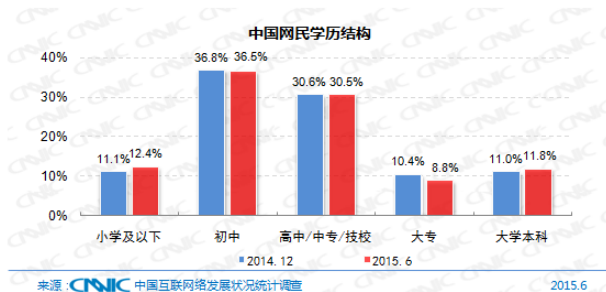


图7 中国网民学历结构图

(四) 职业结构

截至2015年6月,网民中学生群体的占比最高,为24.6%,其次为个体户、自由职业者,比例为22.3%,企业、公司的管理人员和一般职员占比合计达到16.3%,见图8。

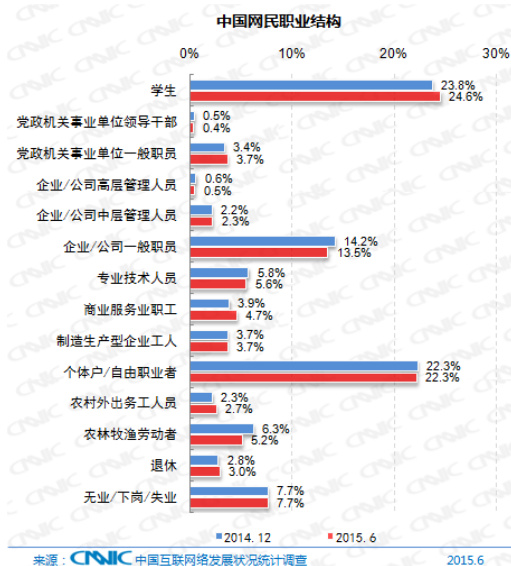


图8 中国网民职业结构图

（五）收入结构

截至 2015 年 6 月，网民中月收入在 2 001 ~ 3 000、3 001 ~ 5 000 元的群体占比最高，分别为 21.0%和 22.4%，见图 9。

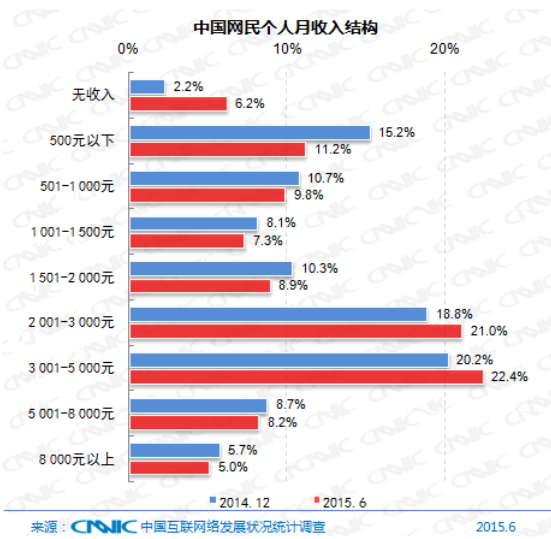


图 9 中国网民个人月收入结构图
（资料来源：第 36 次中国互联网络发展状况统计报告 中国互联网络信息中心 2015-07-23）

总量指标和相对指标

学习要点

- 总量指标和相对指标的意义与分类。
- 相对指标的计算方法。
- 总量指标与相对指标的正确运用。

统计分析是统计整理的继续。在统计分析的各种方法中，综合指标法是最基本的方法。综合指标有总量指标、相对指标和平均指标 3 种基本类型。总量指标反映客观现象总体的基本数量特征，是统计分析的基本统计指标。相对指标在总量指标的基础上计算形成，可用于研究客观现象的数量对比关系。例如，根据国家统计局统计，2014 年年末，我国内地总人口（包括 31 个省、自治区、直辖市和中国人民解放军现役军人，不包括香港、澳门特别行政区和台湾省及海外华侨人数）为 136 782 万人，比上年年末增加了 710 万人。2014 年我国城镇人口比重达到 54.77%，与上年相比，上升 1.04 个百分点。在这里，正是因为利用了总量指标和相对指标对总体的描述，才使我们对我国人口情况有了客观具体的了解。在生活中，媒体上也常辟有“数字新闻”和“新闻调查”等专题栏目，描述被研究对象的特征，使我们对现象有深刻、全面的认识。本章将全面介绍总量指标和相对指标的意义、计算方法及在社会经济生活中的应用。

4.1 总量指标

4.1.1 总量指标的意义与种类

1. 总量指标的意义

总量指标是以绝对数形式反映一定客观现象总体在具体时间和地点条件下的总规模、总水平或工作总量的统计指标，又称统计绝对数（absolute number）。例如，全国人口总数可以反映全国人口总规模，某市某年地区生产总值反映该城市全年三次产业增加值总量。总量指标也就是前面所述的数量指标，总量指标是最基本的综合指标。

在社会经济统计中，总量指标具有重要作用。它可以通过人口总数、劳动力数量、国内生产总值、各种主要产品产量等总量指标，反映一个国家的基本国情和国力；可以为我们编制长

期规划、进行监督管理提供依据。总量指标是综合指标的基本形式，它是形成相对指标和平均指标的必备基础，因为相对指标和平均指标都是总量指标的派生指标。因此，总量指标的计算正确与否直接影响到其他指标计算的正确性。

2. 总量指标的种类

(1) 按反映的内容不同，总量指标可分为总体单位总量和总体标志总量。总体单位总量表示的是一个总体内所包含的总体单位总数，即总体本身规模的大小。总体标志总量是总体各单位某种数量标志值的总和，是说明总体特征的总数量。例如，对某城市居民的粮食消费情况进行研究时，该城市居住的人口总数便是总体单位总量指标，居民的粮食消费总数则是总体标志总量指标。

总体单位总量和总体标志总量并不是一成不变的，随着研究目的和研究对象的不同而变化。例如，当以全国为研究总体来观察某城市的人口数时，上述“人口数”指标就是总体标志总量指标；当研究目的为某市的粮食销售价格时，“粮食消费总数”便是总体单位总量指标。

(2) 按反映的时间状况不同，总量指标可分为时期指标和时点指标。时期指标是反映客观现象总体在一段时期内活动过程的总数量，即流量指标，如产品产量、原材料消耗量、一定时期的基本建设投资额、商品流转额及出生人口数等。这一指标说明时期现象在相应时期内发生的总量是多少。这种现象是连续不断地发生的，每个时期的累计数就表明现象在该时期整个活动过程的总成果。

时点指标是反映客观现象总体在某一时点（瞬间）上状况的总量指标，即存量指标，如期初或期末固定资产额、原材料库存量及人口总数等。这一指标说明时点现象在相应时点上所达到的总量水平的大小。

时期指标和时点指标都是总量指标，这是它们的共同点，但它们又有明显的区别。首先，不同时期的时期指标可以相加，而不同时间点的时点指标不能相加。例如，公司1月份产量100台，二月份产量110台，则我们可以说2个月产量总共210台。若公司1月底员工人数100人，2月底员工人数110人，我们不能说2个月公司共有员工210人。其次，时期指标数值的大小与时间长短有直接关系，而时点指标的大小与时点间的间隔长短无直接关系。我们可以肯定，时间越长产量越多，公司一年的产量必然大于（至少等于）半年的产量。但我们不能肯定时间经过一年时（即年底）员工人数一定大于或小于时间经过半年时（即年中）的员工人数。最后，时期指标的资料搜集可以通过经常性调查取得，而时点指标的资料搜集只能通过一次性调查取得。

4.1.2 总量指标的计量单位

总量指标是社会经济现象总量的具体表现，采用绝对数形式，必须有一定的计量单位。计量的准确与否直接影响总量指标反映客观现象的准确程度。根据事物的性质特点和研究任务不同，总量指标的计量单位可分为下列3种类型。

1. 实物单位

实物单位是根据事物的自然属性和特点所采用的计量单位，如人口总数以“人”进行计量、钢铁产量以“千克”进行计量等。它可以明显而又准确地反映事物的使用价值量的大小，但缺乏综合性。实物单位可进一步分为以下几种。

（1）自然单位。它是根据被研究现象的自然状况来统计其数量的一种计量单位，如职工人数以“人”为单位、汽车以“辆”为单位、鞋以“双”为单位等。

（2）度量衡单位。它是按照统一的度量衡制度的规定来度量客观事物数量的一种计量单位。一般采用国际公制度量衡单位进行计量。例如，煤炭产量以“千克”、布产量以“米”、天然气产量以“立方米”为计量单位，等等。采用度量衡单位的原因是有些现象无法采用自然计量单位来表明其数量，如钢铁、粮食等。另外，有些现象虽然可以采用自然计量单位，但不如度量衡单位准确。统一度量衡制度是准确反映客观事物数量的前提。

（3）双重和复合单位。它是综合两种或两种以上实物单位来完全反映事物实际数量或实际能力的计量单位。双重单位常常以相除的方法将两个单位结合在一起，如电动机功率以“千瓦/台”为计量单位。复合单位则是用相乘的方法将两种计量单位有机地结合在一起来表示事物的数量，如国内旅游出游人数以“人次”为计算单位、旅客运输周转量以“人公里”为计量单位等。

（4）标准实物单位。它是对名称和用途基本相同，但效能不同的事物，按照其物理、化学或其他性质折算成标准实物的计量单位。例如，将发热量不同的煤折合为每千克发热量为29.307 kJ的标准煤计算产量，各种不同含量的化肥用折纯法折算为100%含量的化肥进行计算，集装箱吞吐量按20英尺货柜（TEU twenty foot equivalent unit）为标准箱进行折算，等等。

用实物单位计算的总量指标，称为实物指标。实物指标的优点是能直接反映商品的使用价值和现象的具体内容，因而能直接表明事物的规模水平，但缺点是综合性能差，性质不同的计量单位无法进行汇总。

2. 价值单位

价值单位是采用货币作为价值尺度来度量劳动成果或社会财富的一种计量单位，如工资总额、增加值、地区生产总值、社会商品零售额、产品进出口总额等。

用价值单位计算的总量指标，称为价值指标。价值指标的优点是使不能直接相加的产品总量或商品数量过渡到能够加总，可以综合说明具有不同使用价值的产品总量或商品销售量等的规模或总水平。因此，价值指标具有广泛的综合性和概括能力。其缺点是价值指标脱离了物质内容，比较抽象。

3. 劳动量单位

劳动量单位是采用劳动时间来度量劳动成果的一种计量单位，如出勤工日、定额工时、实际工日等。劳动量指标是评价劳动时间利用程度和计算劳动生产率的依据，同时也是企业编制生产计划和检查生产计划的依据。

4.1.3 总量指标的计算和应用

1. 总量指标的计算

总量指标的计算方法有直接计量法、推算法和估算法。

直接计量法是指对研究对象进行直接记数、测量后，将总量指标的数值计算出来的方法。这种方法要求对总体的所有单位都进行登记，并汇总出所需要的资料，如商品流转统计报表中的库存量、人口普查资料等。

在总量指标不能直接计算或不必要直接计算的情况下，可采用推算法和估算法。常用的推算方法如下。

(1) 因素关系推算法。即利用社会经济现象的各个影响因素之间的关系，根据已知因素来推算未知因素的方法。例如，“销售额=单价×销售量”，这一关系式中的某两项已知就可推算另一未知因素的数值。

(2) 比例关系推算法。即利用各种相关资料的比例关系进行推算的一种方法。例如，某超市零售总额中，团体消费品零售额历年均占零售总额的10%，已知2015年该超市零售总额为8 000万元，利用比例推算法，则该超市2015年团体消费品零售额为800万元（即8 000万元×10%）。

(3) 平衡关系推算法。即利用各种平衡关系来推算未知指标的方法。例如，“期初库存+本期购进=本期销售+期末库存”，这一关系式中的三项已知就可以利用平衡关系推算另一未知因素的数值。

估算法是指运用抽样推断的方法估算总量指标，具体内容将在第6章中介绍。

2. 总量指标的应用

在实际统计工作中，经常运用统计调查的资料汇总计算总量指标，此时，应注意以下问题。

(1) 科学确定总量指标的含义、范围和计算方法。总量指标的计算并非单纯的汇总技术问题。例如，计算工业企业增加值时，首先应对“工业企业”含义加以确切界定，然后是增加值包括的范围和计算方法问题，确定后才能进行正确的统计。

(2) 计算实物总量指标时，要注意现象的同类性。实物指标的同类性是由产品的使用价值决定的，只有使用价值相同的产品才能汇总计算总量指标，把不同使用价值的产品产量简单加总是没有意义的。例如，简单地把煤炭、石油、粮食等产品产量相加就没有确切的经济含义。但是对现象同类性的要求不能绝对化。例如，在计算粮食产量时，就不必考虑粮食是小麦、水稻，还是玉米、高粱。再如，计算货物运输量总量时，产品的同类性就不成为计算的条件，因为它只要求通过货物的重量和里程来计算货物量和货物周转量。

(3) 必须使用统一计量单位。总量指标的计量单位必须按照国家统一的计量单位进行计量。例如，在国际贸易、市场营销等经济管理工作中，为了准确地反映客观事物的总量，在进行各个时期、各个地区甚至国际间资料的对比时，应规定各种统一标准计量单位和统一度



量衡制度。只有这样，才不会造成统计计量方面的差错或混乱，才能客观地统计社会经济现象总体的数量。

我国从1991年起统一使用国际单位制为基础的法定计量单位制度，促进了实物总量指标的准确统计。

(4) 总量指标在实际应用时一般不单独使用，一是可以与同类总量指标相加减，例如，本月产量100万台，比去年同期增长了16万台，截至本月份累计生产380万台；二是与相对指标和平均指标结合应用，例如，本月产量100万台，比上月增长了20%，工人平均月产量达到300台。

4.2 相对指标

4.2.1 相对指标的意义与种类

总量指标虽然可以综合反映社会经济现象的规模、水平和工作总量，但由于现象总体的复杂性，仅根据总量指标难以对客观事物做出正确的判断。相对指标是在总量指标的基础上进行对比而产生的统计分析指标，它有利于反映现象之间的联系状况。

1. 相对指标的意义

相对指标又称统计相对数(relative number)，它是指两个相互联系的统计指标之比。例如，人口的性别比例、人口的出生率和死亡率、资金利润率、经济发展速度、人均国内生产总值、失业率、物价指数等都是相对指标。相对指标的特点在于指标数值的抽象化，就是说，它是把两个具体数值加以概括化所形成的抽象化的数值。通过这种抽象化，使人们对客观现象的数量特征有一个更清晰的认识。

在社会经济生活、国民经济管理和统计研究中，相对指标应用很广泛，具有十分重要的意义。首先，相对指标能够反映社会经济现象之间的数量对比关系，因为发展规模的大小、计划执行的好坏、变化速度的快慢、各种比例协调与否，只有通过比较才能反映出来。其次，相对指标把社会经济现象的绝对差异抽象化，使原来不能直接对比的指标可以进行对比。例如，考察不同地区居民生活的富裕程度时，由于各地区客观条件不同，不能用总量指标直接对比，但如果都以各自的食物支出总额和消费支出总额指标作为依据，计算结构相对指标恩格尔系数，就可以进行比较和深入分析。

2. 相对指标的表现形式

相对指标的表现形式有两种：一种是无名数，另一种是有名数。无名数是一种抽象化的数值，多用倍数或系数、成数、百分数和千分数来表示。

(1) 倍数和系数。它们是将对比基数抽象化为1来计算的相对数。当分子数值比分母数值大很多时，常用倍数表示。当分子数值和分母数值差别不大时，常用系数表示。系数可以大于1，也可以小于1。

(2) 成数。它是将对比基数抽象化为 10 来计算的相对数。例如,粮食产量增加两成,即增长十分之二。

(3) 百分数和千分数。百分数(%)是将对比基数抽象化为 100 来计算的相对数,是相对指标中最常用的表现形式。百分点是百分数的另一种表述形式,它以 1% 为单位,即 1 个百分点等于 1%。它在两个百分数相减的情况下应用。例如,某银行人民币一年期定期存款储蓄利率为 2.5%,若现在下调 0.25 个百分点,则说明现在储蓄利率为 2.25%。

千分数(‰)是将对比基数抽象化为 1 000 而计算的相对数。一般在两个数值对比时,当分子比分母的数值小很多时,则用千分数表示。例如,我国 2014 年人口出生率为 12.37‰;死亡人口 977 万人,死亡率为 7.16‰;自然增长率为 5.21‰。

(4) 翻番数。翻番数是指两个相比较的数值中,一个数是另一个数的“2”倍,则 n 是番数。例如,某集团 2015 年的工业总产值为 240 亿元,计划 2020 年翻一番,则该集团 2020 年工业总产值应达到 480 亿元;计划翻两番,则该集团 2020 年工业总产值应达到 960 亿元;翻三番则应达到 1 920 亿元。

有名数主要用于计算部分强度相对指标,如人口密度用“人/平方公里”、居民人均生活费收入用“元/人”等表示。通常是将进行对比的两个经济指标的计量单位同时使用,以双重单位表示。

3. 相对指标的种类

相对指标既然是两个有联系的指标之比,那么,哪两种指标之间有联系?把这两种指标进行对比之后能说明什么问题?根据这个标准,有联系的指标主要有以下几种:结构相对指标、比例相对指标、比较相对指标、动态相对指标、强度相对指标和计划完成程度相对指标。这些指标从不同方面表明现象的相对水平,表明现象的结构组成、发展程度、比例关系等。

4.2.2 相对指标的计算

随着研究目的和任务不同,对比的基数也不同,相对指标就会产生不同的种类和计算方法,以下介绍常用的几种。

1. 结构相对指标

结构相对指标是指在分组的基础上,将总体区分为性质不同的各个部分,以各部分指标数值与总体全部数值相比所得的相对数。它反映现象各部分在总体中所占的比重或比率,说明现象内部的组成状况与数量关系,其计算公式为:

$$\text{结构相对指标} = \frac{\text{总体中某一部分数值}}{\text{总体全部数值}} \times 100\% \quad (4-1)$$

例如,根据 2010 年全国第六次人口普查数据公报,我国普查登记的大陆 31 个省、自治区、直辖市和现役军人的人口共 1 339 724 852 人,其中男性人口为 686 852 572 人,占 51.27%;女性人口为 652 872 280 人,占 48.73%。这表明了我国人口的性别构成状况。

结构相对指标一般以百分数的形式表示。因为总体的全部数值等于总体内部各部分数值之和，所以总体内各部分占总体的比重（百分比）之和等于 100%或 1。应注意，结构相对指标的分子分母位置不能互换。

结构相对指标在统计分析中应用非常广泛，常应用于消费结构分析。消费结构是指各类消费支出在总消费支出中所占的比重。19 世纪德国统计学家恩格尔通过对英国、法国、德国、比利时等国居民家庭收支进行分析研究，提出了恩格尔定律，即随着家庭收入的增加，家庭收入或总支出中用于食品方面的支出比重越来越小。反映这个定律的结构相对数，称为恩格尔系数。恩格尔系数=食品支出总额/消费支出总额。

将不同时期的结构相对指标进行对比，可以通过总体结构的变动，观察客观事物变化的进程。一般来说，事物的变化总是先从内部结构演变开始的，这种变化常常反映着事物发展由量变到质变的过程。掌握这一进程并加以分析，有助于我们认识事物发展的规律性。

【实例 4.1】 2010—2014 年我国城镇和乡村人口数比重资料如表 4.1 所示。

表 4.1 我国 2010—2014 年城乡人口数比重

指 标	2010	2011	2012	2013	2014
城镇人口数比重（%）	49.95	51.27	52.57	53.73	54.77
乡村人口数比重（%）	50.05	48.73	47.43	46.27	45.23

（资料来源：国家统计局）

从表 4.1 的数据可以看出，2010—2014 年我国城镇人口数比重逐年上升，乡村人口数比重逐年下降，城镇化进程正在逐渐加快。

2. 比例相对指标

比例相对指标是同一总体内不同组成部分之间的比值，反映各组成部分之间的联系和比例关系，其计算公式为：

$$\text{比例相对指标} = \frac{\text{总体中某部分数值}}{\text{总体中另一部分数值}} \times 100\%$$

(4-2)

比例相对指标可以用百分数表示，也可以用一比几或几比几的形式表示。比例相对指标的分子、分母同属一个总体，而且分子、分母的位置可以互换。

【实例 4.2】 以 2010 年 11 月 1 日零时为标准时点，我国大陆 31 个省、自治区、直辖市和现役军人共 1 339 724 852 人，其中居住在城镇的人口为 665 575 306 人，居住在乡村的人口为 674 149 546 人，则城镇人口与乡村人口的比例可表示为 1：1.013。

【实例 4.3】 我国 2014 年国内生产总值 636 463 亿元。其中，第一产业增加值 58 332 亿元，第二产业增加值 271 392 亿元，第三产业增加值 306 739 亿元。第一、第二和第三产业增加值的比例为 9.2：42.6：48.2。

分析总体中若干部分的比例关系时可采用连比形式（1：*m*：*n*）。为了能清楚地反映各部分之间的数量关系，连比组数不宜过多。

比例相对指标对于国民经济宏观调控具有重要意义。例如，人口性别比例、工农业总值中农轻重比例、积累与消费比例等，利用比例相对指标可以分析国民经济中的各种比例关系，从而调整不合理的比例，促使市场经济稳步协调发展。

3. 比较相对指标

比较相对指标又称类比相对指标，是将两个同类指标作静态对比得出的相对指标，表明同类现象在不同空间（如在各 国、各地、各单位）下的数量对比关系，其计算公式为：

$$\text{比例相对指标} = \frac{\text{某条件下某类指标数值}}{\text{另一条件下同类指标数值}} \times 100\%$$

(4-3)

式中，分子与分母现象所属统计指标的含义、口径、计算方法和计量单位必须一致。比较相对指标一般用百分数或倍数表示。

【实例 4.4】 2014 年北京、上海两城市有关资料如表 4.2 所示。

表 4.2 2014 年北京、上海两城市相关资料比较

城 市	年末常住人口（万人）	地区生产总值（亿元）	人均地区生产总值（按常住人口 万元/人）
北京	2 151.6	21 330.8	10.0
上海	2 425.68	23 560.94	9.73
比较相对指标（以北京为 100）	112.74	110.46	97.3

从表 4.2 的数据可看出，虽然上海市地区生产总值比北京市多 10.46%，但由于上海的常住人口数比北京市多 12.74%，因此上海的人均地区生产总值比北京市少 2.7%。

在实际工作中，计算比较相对指标时，作为比较基数的分母除了可以取常规的与分子分属不同空间的同类指标外，还可以将比较的基数典型化。例如，将本单位产品的质量、成本、单耗等各项技术经济指标都与国内外同行业的先进水平对比，与国家规定的质量标准进行对比等，从而找出差距，为提高本单位生产管理水平提供依据。

比较相对指标与比例相对指标类似，分子与分母也可以互换。两者的差别在于：比例相对指标是同一总体的不同部分比较，而比较相对指标是同类指标的不同空间比较。比例相对指标反映的比例关系有时是一种客观标准，违背这个标准，就会造成比例关系的失调；比较相对指标反映的是事物之间的对比关系，一般不存在比例正常或失调的关系。

4. 动态相对指标

动态相对指标又称为动态相对数或发展速度，是同类事物在不同时间上的指标之比，反映现象在时间上的发展变化方向 and 变化程度。其计算公式为：

$$\text{动态相对指标} = \frac{\text{报告期指标数值}}{\text{基期指标数值}} \times 100\%$$

(4-4)

一般把用来作为对比基础的时期称为基期，而把同基期对比的时期称为报告期或计算期。



【实例 4.5】 以 2010 年 11 月 1 日零时为标准时点，我国大陆 31 个省、自治区、直辖市和现役军人共 1 339 724 852 人，而第五次全国人口普查 2000 年 11 月 1 日零时共有 1 265 825 048 人，则

$$\text{动态相对指标} = \frac{1\,339\,724\,852}{1\,265\,825\,048} \times 100\% = 105.84\%$$

这表明十年间我国大陆人口共增长 5.84%。它反映了我国大陆人口在十年间的变化情况。

动态相对指标将在本书第 7 章时间数列中详细介绍。

5. 强度相对指标

强度相对指标，又称为强度相对数，是两个性质不同但有一定联系的总量指标对比的结果，用来表明现象的强度、密度和普通程度的综合指标。其计算公式为：

$$\text{强度相对指标} = \frac{\text{某一总量指标数值}}{\text{另一有联系而性质不同的总量指标数值}} \quad (4-5)$$

强度相对指标一般采用有名数形式。例如，我国的人均粮食产量、人均 GDP、居民人均可支配收入及人均汽车拥有量等，采用的是“千克/人”、“元/人”、“辆/千人”等复名数形式；商品流转速度用“次”或“天”表示，采用的是单名数形式。但也有采用无名数形式的，如人口自然增长率、商品流通费率、企业资金利润率等采用千分数、百分数等无名数形式。

【实例 4.6】 2014 年我国年平均人口数为 13.7 亿人，我国国土总面积为 960 万平方公里，计算强度相对指标。

解：我国人口密度 = $137\,000 / 960 = 142.71$ （人/平方公里）

强度相对指标常带有“平均”的含义，但由于它的分子、分母分属两个不同总体，所以它与平均指标不同。同时，有些强度相对指标的分子与分母可以互换，因此，它有正指标和逆指标之分。强度相对指标数值的大小与现象的强度、密度和普遍程度成正比的是正指标；强度相对指标数值的大小与现象的强度、密度和普遍程度成反比的是逆指标。例如，在医疗卫生统计中，每千人拥有的医生数是正指标，每个医生所服务的人口数是逆指标。一般说来，正指标越大越好，逆指标则越小越好。

但要注意的是，在实际工作中，并不是所有的强度相对指标都有正指标和逆指标之分。例如，人口出生率、人口死亡率等指标的分子、分母是不能互换的。因此，在计算和应用强度相对指标时，应根据研究目的、使用习惯及说明问题的难易程度来选择使用正指标或逆指标。

6. 计划完成程度相对指标

计划完成程度相对指标是实际完成数与相应的计划数之比，用以表明计划完成的程度，也叫计划完成相对数或计划完成百分比（percentage of plan fulfilment）。计划完成程度相对指标是检查、监督计划执行情况的重要指标，也是正确评价工作成绩的重要依据。

计划完成程度相对指标一般用百分数表示，其基本计算公式为：

$$\text{计划完成程度相对指标} = \frac{\text{实际完成数}}{\text{计划数}} \times 100\% \quad (4-6)$$

式中,分母是原定的计划指标,分子则是计划执行过程中或执行过程结束后统计出来的实际完成数值。一般要求分子与分母在指标含义、计算口径、计算方法、计量单位、空间范围等方面一致。但实际完成数与计划数所包含的时期长短可以是相同的,也可以不同,两者对比的意义有所区别。

(1) 计划完成程度的计算。在制订计划时,由于具体情况与计算要求不同,计划任务数的表现形式有绝对数、相对数、平均数三种。因此计算计划完成程度指标时,既可以用绝对数直接对比求得,也可以用相对数和平均数对比求得,具体计算方法如下。

① 计划指标为总量指标时计算计划完成程度相对指标。这是计算计划完成程度指标的基本方法,具体计算公式为:

$$\text{计划完成程度相对指标} = \frac{\text{实际水平}}{\text{计划水平}} \times 100\% \quad (4-7)$$

【实例 4.7】 某销售部门 2015 年上半年商品销售额计划为 2 000 万元,实际完成 2 100 万元,求该销售部门上半年商品销售额计划完成情况。

$$\text{解: 计划完成程度} = \frac{2\,100}{2\,000} \times 100\% = 105\%$$

计算结果表明,该销售部门 2015 年上半年商品销售额计划完成程度为 105%,即超额完成 5% (105%-1),超额完成 100 万元 (2 100 万元-2 000 万元)。

② 计划指标为相对指标时计算计划完成程度相对指标。在实际经济工作中,有时计划指标是以相对指标形式规定的,即用提高或降低百分比来规定计划任务,如劳动生产率、成本降低率、流通费用率等。在这种情况下,计算计划完成程度相对指标就需要以实际完成百分比与计划规定百分比进行对比,其计算公式为:

$$\begin{aligned} \text{计划完成程度相对指标} &= \frac{\text{实际完成百分比}}{\text{计划完成百分比}} \times 100\% \\ &= \frac{\text{实际水平} / \text{基期水平}}{\text{计划水平} / \text{基期水平}} \times 100\% \\ &= \frac{\text{实际水平}}{\text{计划水平}} \times 100\% \end{aligned}$$

式中,计划规定百分比通常是以基期(如上年度)实际水平为基础确定的,也就是说,在计划规定计划任务时,将计划水平与基期实际水平对比来求得计划提高或降低的百分比;实际完成百分比也以基期水平为基础,即以报告期实际水平与基期水平对比来求得实际提高或降低的百分比。因此,上述公式可变形为:

$$\text{计划完成程度相对指标} = \frac{1 \pm \text{实际提高(降低)百分比}}{1 \pm \text{计划提高(降低)百分比}} \times 100\% \quad (4-8)$$

可见，用总量指标和相对指标计算出来的指标结果是一致的。式（4-8）还表明：利用相对指标计算计划完成程度的指标时，不能用实际增长率（降低率）除以计划增长率（降低率）来求得，而应包含原有基础（即以基期水平为100%）在内，这样计算才符合公式含义，计算结果才能与用总量指标计算出来的结果相一致。

【实例4.8】 某鞋业公司计划规定劳动生产率2015年比2014年提高10%，实际提高了15%，求劳动生产率计划完成情况。

解：

$$\text{劳动生产率计划完成长度} = \frac{1+15\%}{1+10\%} \times 100\% = \frac{115\%}{110\%} \times 100\% = 104.5\%$$

计算结果表明，该企业劳动生产率得到提高，超额4.5%完成计划。

【实例4.9】 某鞋业公司产品成本计划规定2015年比2014年降低4%，实际降低6%，求该企业成本降低率计划完成情况。

解：

$$\text{成本降低率计划完成程度} = \frac{1-6\%}{1-4\%} \times 100\% = \frac{94\%}{96\%} \times 100\% = 97.92\%$$

计算结果显示，计划完成百分比小于百分之百，因为它是成本降低计划，所以这同样

表明该企业完成了计划，而且该企业产品实际成本比计划规定多降低2.08%（100%–97.92%）。

注意，在实际工作中，有时也用百分点描述计划完成情况。例如，在【实例4.8】中，劳动生产率实际比计划多提高了5个百分点（15%–10%=5%），但不能说成提高5%；在【实例4.9】中，公司的产品成本实际比计划降低了2个百分点（6%–4%=2%），但不能说成多降低2%。百分点是指以百分数形式表示的相对数指标的增减变动幅度或对比差额，是指被比较的相对数指标之间的增减量，而不是它们之间的比值。

③ 计划指标为平均指标时计算计划完成程度相对指标。如果计划指标以平均指标的形式提出，如产品单位成本、员工平均工资等，这时可直接运用计划完成程度相对指标的基本公式，以实际平均水平与计划平均水平进行对比计算。

【实例4.10】 2015年红星自行车有限公司生产的自行车，产品单位成本计划为500元，实际为470元，求自行车产品单位成本计划完成情况。

解：

$$\text{单位成本计划完成程度} = \frac{471}{500} \times 100\% = 94\%$$

计算结果表明，红星自行车有限公司自行车产品单位成本实际比计划降低了6%，单位成本降低了30元。

(2) 计算和分析长期计划完成情况的方法。由于长期计划（如 5 年计划）中所规定的指标性质不同，所以表示方法也不同，因此产生了两种不完全相同的考核长期计划执行情况的方法。

① 水平法。如果在长期计划中，只规定计划期最末一年应达到的水平，则在计划期中，哪一期达到了计划数，则该期就完成了计划。用水平法检查长期计划执行情况的计算公式为：

$$\text{计划完成程度} = \frac{\text{长期计划末年实际达到的水平}}{\text{长期计划规定的末年水平}} \times 100\% \tag{4-9}$$

【实例 4.11】 某公司“十二五”计划规定，公司在 2015 年应达到年产量 200 万吨水平，各年实际生产情况如表 4.3 所示，检查该公司产量“十二五”计划完成情况。

表 4.3 某公司“十二五”期间实际产量 单位：万吨

时间	2011 年	2012 年	2013 年	2014 年				2015 年			
				第一季度	第二季度	第三季度	第四季度	第一季度	第二季度	第三季度	第四季度
产量	143	159	168	35	40	45	55	60	62	68	70

解：

$$\text{2015 年公司产量计划完成程度} = \frac{60 + 62 + 68 + 70}{200} = \frac{260}{200} \times 100\% = 130\%$$

计算结果表明，该公司产量超额 30% 完成了“十二五”计划。那么，提前完成计划时间是多少呢？

公司在 2011 年、2012 年、2013 年、2014 年的产量均未达到 200 万吨，而从 2014 年的第二季度到 2015 年的第一季度的产量刚好达到 200 万吨（40+45+55+60=200），说明到 2015 年的第一季度末完成了计划，提前了 3 个季度。

利用水平法考核长期计划执行情况，计算提前完成任务的时间时，是以“统计年”计算的。即只要有连续一年时间（可以跨年度）实际完成水平达到最后一年计划水平，就算完成了长期计划，余下的时间就是提前完成计划时间。

② 累计法。如果在长期计划中，计划指标规定的是整个计划期内累计应达到的水平，则采用累计法考核长期计划执行情况，即将计划期内各年实际完成累计数与长期计划规定的累计数进行对比来检查计划完成情况。其计算公式为：

$$\text{计划完成程度} = \frac{\text{长期计划期间实际累计完成数}}{\text{长期计划规定的累计数}} \times 100\% \tag{4-10}$$

【实例 4.12】 某地区“十二五”计划规定 5 年内基本建设投资总额为 30 亿元，实际执行结果如表 4.4 所示，检查该地区基本建设投资总额“十二五”计划完成情况。

表 4.4 某地区“十二五”计划期间基本建设投资总额

单位：亿元

时 间	2011 年	2012 年	2013 年	2014 年		2015 年			
				上半年	下半年	第一季度	第二季度	第三季度	第四季度
基本建设投资额	4.5	5.8	6.7	3.8	4.6	2.4	2.2	2.3	2.5

解：

该地区基本建设投资总额计划完成程度= $\frac{34.8}{30} \times 100\% = 116\%$

结果表明，该地区基本建设投资额超额 16%完成了“十二五”计划。那么，提前了多少时间完成计划呢？

该地区基本建设投资额从 2011 年起至 2015 年第二季度止，实际累计完成了 30 亿元（4.5+5.8+6.7+3.8+4.6+2.4+2.2=30），说明该地区提前两个季度完成了基本建设投资额的“十二五”计划。

利用累计法考核长期计划执行情况时，计算提前完成计划的时间是很有意义的。将计划期的全部时间减去自计划执行之日起至累计实际完成数已达到计划指标的日期止，其剩余的时间即为提前完成计划的时间。

（3）计划执行进度的检查。在实际应用中，利用计划完成程度指标检查计划执行情况时，往往有两种情况。一种是检查计划执行的结果，带有总结的性质，这时公式中实际水平与计划水平所包含的时期长短是完全一致的；另一种情况是检查计划执行的进度和均衡程度，这时公式中实际水平与计划水平所包含的时期长短是不一样的。

【实例 4.13】 某公司所属三个企业 2015 年全年和第三季度计划产值与实际完成产值的资料如表 4.5 所示。

表 4.5 某公司三个企业产值完成情况资料

企 业	计划总产值		实际完成总产值		第三季度总产值计 划完成程度（%）	到第三季度止年总产 值计划完成程度（%）
	全年	其中第三季度	第三季度	累计到第三季度止		
	（1）	（2）	（3）	（4）	（5）=（3）/（2）	（6）=（4）/（1）
甲	5 000	1 250	1 320	3 800	105.6	76.0
乙	1 800	480	480	1 349	100.0	74.9
丙	700	175	168	476	96.0	68.0
合计或平均	7 500	1 905	1 968	5 625	103.3	75.0

在表 4.5 中，第（5）、（6）栏中都是计划完成程度指标，但各自说明不同的问题。第（5）栏说明第三季度产值计划的完成情况：全公司第三季度超额完成计划 3.3%；所属各企业完成程度不同，甲企业超额完成 5.6%，丙企业则未完成计划。第（6）栏说明到第三季度止，实际完成全年计划产值的进度情况：全公司完成 75%，正好是全年计划的 3/4，符合稳定生产的要求；但各企业的完成情况并不均衡，丙企业未实现进度要求，只完成全年计划的 68%，可见丙企业产值计划能否完成，是能否实现全公司全年计划的关键。

由以上分析可知,计划完成程度指标不仅可以检查计划执行的结果,更重要的是可以检查和分析计划执行的进度,反映计划执行的均衡性。

4.2.3 相对指标的应用

以上所叙述的各种相对指标,从不同角度反映了现象之间的联系,要正确计算和运用这些相对指标,发挥它们各自的作用,在统计实际中还应注意以下问题。

1. 对比指标的可比性原则

相对指标是通过相互联系的统计指标之间的对比来研究和分析客观现象之间的数量对比关系的,因此保持对比指标之间的可比性,是运用相对指标的基本原则。客观内容不可比、统计范围或者统计口径不可比、计算方法和计量单位不可比,均会导致统计分析结论的谬误。

在实际工作中,可比性是计算和运用相对指标的重要原则。可比性主要包括以下两个方面:

(1) 总体范围可比。例如,在新的一年里,某公司兼并了同行业另一公司,企业规模扩大了,若计算产量动态相对指标,由于总体范围的变化,两年产量不可直接对比。另外,相对指标的分子与分母在范围上要相互适应。例如,文盲和半文盲占总人口的比例,按规定文盲和半文盲是指12周岁以上不识字或识字很少的人,这样在分母的总人口中,就应将12周岁以下的人口扣除掉才比较合理。

(2) 计量单位和计算方法可比。这是将同类现象在不同空间和时间上进行对比时应注意的问题。例如,对比两个不同时期的地区总产值,就要考虑是按现行价格计算还是按不变价格计算。因为只有扣除价格变动因素之后,才能真正比较出不同时期产值的变化。如果是在国际间进行两个指标的对比,就更应慎重考虑指标的范围、内容和计算方法是否具有可比性。例如,西方国家的国民收入与我国国民收入指标的经济含义、所包括的范围及计算方法都有所不同,不能直接进行对比。

遇到分子与分母不可比时,可以进行调整以使它们可比。例如,不同时期的行政区划、组织机构或隶属关系有所不同,这时应以报告期为准,对基期数值进行调整。又如,指标计算方法、范围和口径的改变,使指标数值不能直接对比,这时也需将基期数据按报告期口径进行调整,以使两指标变为可比。分子与分母的计量单位不同时,也应先将其化为统一,才可比较。

2. 相对指标与统计分组法的结合运用

统计分组法是统计分析中经常运用的重要方法,结构相对指标、比例相对指标等必须以其为基础加以运用。除此之外,为了全面地分析客观现象的计划完成情况、发展变化关系,同样必须运用科学的分组方法。因此,相对指标与统计分组的科学配合是统计研究的一项重要内容。

3. 各种相互有联系的相对指标的结合运用

在实际工作中，把相互有联系的各种相对指标结合起来运用，可以更深刻、更全面地说明客观事物的性质，使我们得出正确结论。例如，某企业增加值计划完成相对指标为 110%，只能说明它超额完成了计划。如果其增加值和上年比较的动态相对指标为 108%，将两者联系起来，就可以说明它在高速度发展的基础上超额完成了计划。如果再把该企业的品种、质量、原材料消耗、劳动生产率、成本和利润等方面的相对指标联系起来观察，就能对该企业的工作做出全面的正确评价。再如，评价某个工业企业的产品成本水平时，既要看动态升降程度，又要看计划完成情况，还要看与同行先进水平之间的差异程度，这样才能做出客观的分析。

4. 总量指标与相对指标的结合运用

由于总量指标和相对指标的作用不同，为了能获得对社会经济现象比较完整和深入的认识，需要将二者结合起来运用。

相对指标与总量指标相比，能更明显反映现象之间的联系和对比关系。但相对指标一般是由两个总量指标对比形成的，其比值是把总量指标（即绝对数）抽象化之后的结果。因此，仅从相对指标上看不出现象原有的规模和水平，这就掩盖了现象之间绝对量上的差别。由于在事物的比较过程中，有时候会出现总量指标较大、相对指标较小，或总量指标较小、相对指标较大的现象，因此，在运用相对指标研究客观现象的数量对比关系时，必须联系其所依据的基本指标，即必须与其背后的绝对水平和绝对差异相结合，以便客观地评价事物的发展变化情况。例如，2014 年上海市地区生产总值为 23 560.94 亿元，甘肃省为 6 835.27 亿元，若两地区在 2015 年均增加 7%，从相对水平来看增幅相同，但结合绝对水平来看，上海是在基期的 23 560.94 亿元基础上增长 7%，增长的绝对量为 1 649.27 亿元，而甘肃是在基期 6 835.27 亿元的基础上增长 7%，增长的绝对量是 478.47 亿元。从增加的实际效果看，上海是甘肃的 3.45 倍多。所以，只有把相对指标与总量指标结合运用才能更全面地分析和说明问题。

4.3 Excel在总量指标和相对指标中的应用

4.3.1 Excel在总量指标中的应用

Excel 中没有专门的总量指标和相对指标的分析功能，该项功能的实现需要使用者根据总量指标和相对指标的相关知识，结合客观实际资料，编辑公式进行计算。

【实例 4.14】 根据表 4.4 所给出的资料，利用 Excel 计算某公司“十二五”计划期间基本建设实际投资总额。

解：具体的操作步骤如下，结果如图 4.1 所示。

	A	B	C	D	E	F	G	H	I	J
1	时间	2011 年	2012 年	2013 年	2014 年		2015 年			
2					上半年	下半年	第一季度	第二季度	第三季度	第四季度
3	基本建设投资额	4.5	5.8	6.7	3.8	4.6	2.4	2.2	2.3	2.5
4										
5	34.8									

图 4.1 总量指标计算数据图

- 第一步：输入样本数据。
- 第二步：计算 5 年基本建设实际投资总额。在 A5 处输入 “=sum(B3:J3)”，回车得结果。

4.3.2 Excel在相对指标中的应用

【实例 4.15】 根据表 4.4 所给出的资料，利用 Excel 计算某公司“十二五”计划期间基本建设投资计划完成相对数。

解：具体的操作步骤如下，结果如图 4.2 所示。

	A	B	C	D	E	F	G	H	I	J
1	时间	2011 年	2012 年	2013 年	2014 年		2015 年			
2					上半年	下半年	第一季度	第二季度	第三季度	第四季度
3	基本建设投资额	4.5	5.8	6.7	3.8	4.6	2.4	2.2	2.3	2.5
4										
5	1.16									

图 4.2 计划完成相对指标计算数据图

- 第一步：输入样本数据。
- 第二步：计算基本建设投资计划完成相对数。在 A5 处输入 “=(B3+C3+D3+E3+F3+G3+H3+I3+J3)/30*100%”，回车得结果。
- 关于其他相对指标的计算亦是如此。

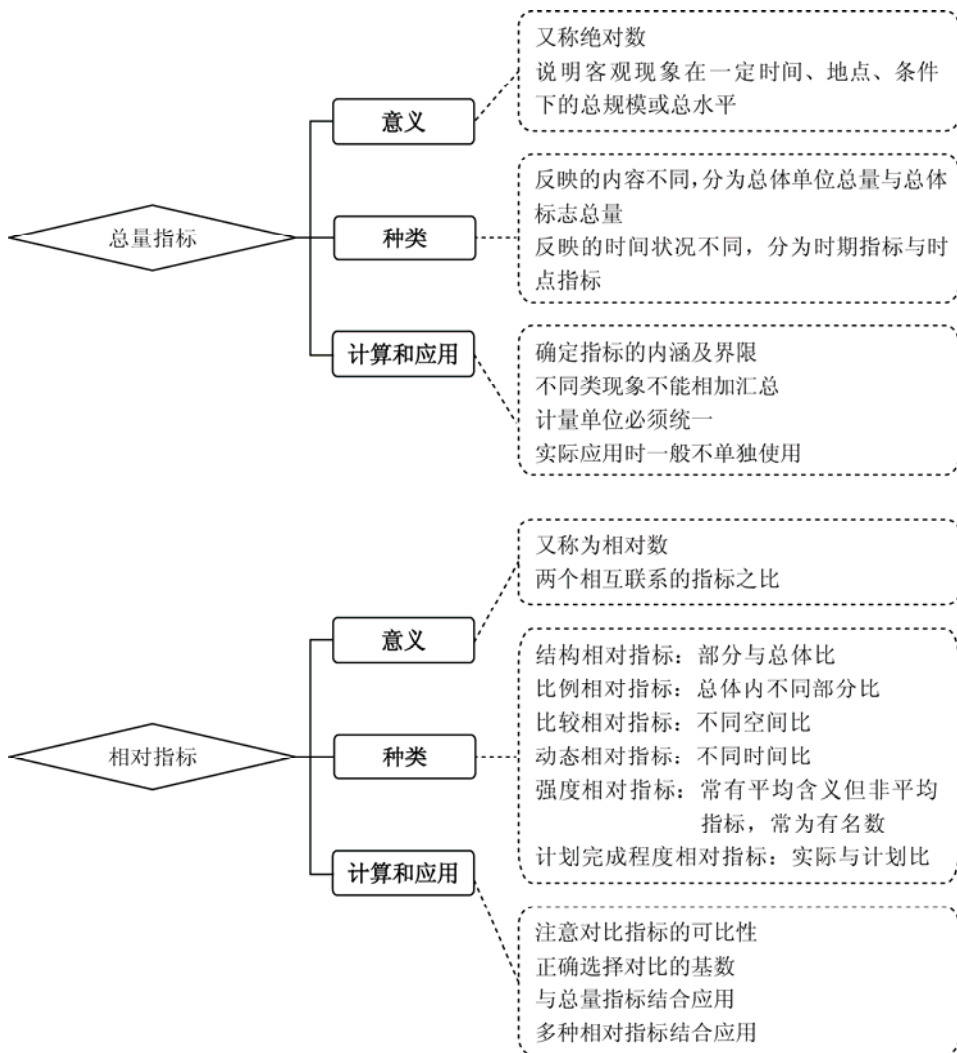


统计术语

统计指标	statistical indicator	绝对数	absolute number
相对数	relative number	计划完成百分比	percentage of plan fulfillment
强度相对指标	severity rate	结构相对数（比重）	proportion



重点知识梳理



习题与实践训练

一、判断题

1. 总量指标按其反映的内容不同分为总体单位总量和总体标志总量。 ()
2. 一个总量指标是总体单位总量还是总体标志总量是固定不变的。 ()
3. 只有当计划完成程度大于 100%时才表示超额完成计划,而小于 100%则表示未完成计划。 ()
4. 强度相对指标有些是用有名数表示的,因此都可计算正指标和逆指标。 ()
5. 企业计划规定,2015 年第一季度的单位产品成本比去年同期降低 15%,实际执行结果降低 7.5%,则企业仅完成单位产品成本计划的一半。 ()

6. 东方电器公司 2015 年第一季度洗衣机产量与春光电器公司同期产量的比率是比例相对指标。 ()
7. 据抽样调查,上海市 2014 年城市居民家庭人均年可支配收入 47 710 元,比上年增长 8.8%,则 8.8%是强度相对指标。 ()
8. 2014 年年末全国共有公共图书馆 3 110 个,文化馆 3 311 个。它们是总体标志总量,同时是时点指标。 ()
9. 比较相对指标既可以用绝对指标对比,也可以用相对指标或平均指标对比。 ()
10. 2014 年我国第一、二、三产业结构比例为 9.2:42.6:48.2,这属于比较相对指标。 ()

二、单项选择题

1. 深圳市 2014 年年末全市常住人口 1 077.89 万人,该数字说明全市人口 ()。
- A. 在年内发展的总规模 B. 在统计时点的总规模
- C. 在年初与年末间隔内发展的总规模 D. 自年初至年末增加的总规模
2. 在出生婴儿中,男性占 53%,女性占 47%,这是 ()。
- A. 结构相对指标 B. 强度相对指标
- C. 比较相对指标 D. 比例相对指标
3. 万华企业生产的变速自行车上年实际成本为 450 元,本年计划降低 4%,实际降低了 5%,则成本降低计划超额完成程度为 ()。
- A. 95% B. 98.9% C. 1% D. 1.04%
4. 按全国人口平均的粮食产量指标是 ()。
- A. 平均指标 B. 强度相对指标
- C. 比较相对指标 D. 结构相对指标
5. 若计划规定年产量比上一年增加 5%,实际增加了 6%,则年产量计划完成 ()。
- A. 120% B. 100.95% C. 101% D. 106%
6. 红星商店计划销售量比去年提高 10%,实际提高 15%,则销售量计划完成程度为 ()。
- A. 150% B. 5% C. 4.5% D. 104.5%
7. 总量指标按反映现象的时间状况不同可以分为 ()。
- A. 时期指标和时点指标
- B. 数量指标和质量指标
- C. 总体单位总量指标和总体标志总量指标
- D. 实物指标和价值指标
8. 下列指标属于总量指标的是 ()。
- A. 出勤率 B. 及格率 C. 人均粮食占有量 D. 学生人数
9. 强生企业计算器产品的年产量为 32 万台,期末库存为 5 万台,这两个指标是 ()。
- A. 时期指标
- B. 时点指标
- C. 前者是时期指标,后者是时点指标
- D. 前者是时点指标,后者是时期指标

10. 某市对所有医院进行调查，其中该市儿童医院调查结果为：共有医生、护士 620 人，其中医生 300 人，护士 320 人；病床数 530 张，医疗设备价值 4 200 万元。上述数值中，总量指标有（ ）。

- A. 1 个 B. 3 个 C. 5 个 D. 0 个

11. 下列指标中属于时点指标的是（ ）。

- A. 商品销售额 B. 商品销售量
C. 平均每人销售额 D. 商品库存额

12. 中网公司 3 月份完成的货运量为 48 000 吨公里。这里的“吨公里”是（ ）单位。

- A. 自然 B. 标准实物 C. 复合 D. 度量衡

13. 某月份甲工厂工人的出勤率属于（ ）。

- A. 结构相对指标 B. 比例相对指标
C. 强度相对指标 D. 相对计划完成指标

14. 人均钢产量属于（ ）。

- A. 平均指标 B. 强度相对指标
C. 比例相对指标 D. 比较相对指标

15. 计划规定成本降低 3%，实际降低了 5%，则计划完成（ ）。

- A. 98.1% B. 102.1% C. 101.9% D. 97.95%

16. 按照计划，现年产量比上年应增加 30%，实际比计划少完成 10%，同上年相比现年产量的实际增长程度为（ ）。

- A. 10% B. 17% C. 20% D. 40%

17. 如果计划任务数是五年计划中规定的最后一年应达到的水平，则计算计划完成程度相对指标可采用（ ）。

- A. 简单平均法 B. 加权平均法 C. 累计法 D. 水平法

18. 某区有 100 个副食品零售商店，商业职工为 2 800 人，商业零售总额为 5 600 万元，在研究商业职工分布和劳动效率的情况时（ ）。

- A. 100 个商店是标志总量指标，又是总体单位数
B. 2 800 人既是标志总量指标，又是总体单位数
C. 5 600 万元既是标志总量指标，又是总体单位数
D. 每个商店的零售额既是标志总量，又是总体单位数

19. 某车间 7 月份在生产老产品的同时，新产品首次小批量投产，出现了 4 件废品，全车间的废品率为 1.3%。8 月份老产品下马，新产品大批投产，全部制品 10 000 件，其中废品 12 件，则 8 月份产品质量（ ）。

- A. 提高 B. 下降 C. 不变 D. 无法确定

20. 在长虹电视机的年末清库中，其年末库存额是（ ）。

- A. 时期指标和实物指标 B. 时期指标和实物指标
C. 时期指标和价值指标 D. 时点指标和价值指标

三、多项选择题

1. 商业银行 2015 年年底的居民储蓄存款额是（ ）。

- A. 综合指标 B. 单位总量指标 C. 标志总量指标
D. 时期指标 E. 时点指标
2. 分子与分母不可互换计算的相对指标是 ()。
- A. 计划完成程度相对指标
B. 动态相对指标
C. 结构相对指标
D. 强度相对指标
E. 比较相对指标
3. 下列哪些是绝对指标 ()。
- A. 某商店月末商品库存额
B. 某地区人口净增加数
C. 全国高等学校历年毕业生人数
D. 某工厂月末在册人数
E. 按人口平均钢产量
4. 实物计量单位包括 ()。
- A. 货币单位 B. 劳动单位 C. 自然单位
D. 度量衡单位 E. 标准实物单位
5. 时期指标的特点是 ()。
- A. 不同时期的指标可以累计
B. 不同时期的指标不可以累计
C. 其数值的大小与其说明的时期长短相关
D. 其数值的大小与其说明的时期长短无关
E. 一般是通过经常性调查得到的
6. 相对指标中由不同总体数值对比的有 ()。
- A. 结构相对指标 B. 强度相对指标 C. 比例相对指标
D. 比较相对指标 E. 动态相对指标
7. 下列统计指标属于强度相对指标的是 ()。
- A. 人口密度 B. 人均国民收入 C. 人口死亡率
D. 农民人均纯收入 E. 经济发展速度
8. 下列属于时点指标的有 ()。
- A. 某地区人口数
B. 某地区人口死亡数
C. 某市在校学生数
D. 某地区基本建设投资额
E. 某农场每年拖拉机台数
9. 按实际汇率计, 2014 年美国 GDP 为 18.287 万亿美元, 中国为 11.285 万亿美元, GDP 这个指标是 ()。
- A. 数量指标 B. 时期指标 C. 时点指标
D. 相对指标 E. 价值指标

10. 下列指标属于比较相对指标的有（ ）。
- A. 男生人数占全班学生总数的 51%
 - B. 中国人口是美国人口的 4.2 倍
 - C. 某商场 9 月份费用额与销售额相比，费用额是销售额的 8%
 - D. 某公司 2015 年产量是 2005 年产量的 1.5 倍
 - E. 2014 年北京市地区生产总值 21 330.8 亿元是深圳市地区生产总值 16 001.98 亿元的 1.33 倍。

四、填空题

1. 总量指标按反映总体内容的不同，可分为_____和_____，按反映时间状况的不同，可分为_____和_____。
2. 总量指标的计算单位除实物单位、货币单位以外，还有_____。
3. 相对指标的表现形式有两种：_____和_____，除强度相对指标用_____表示外，其他都用_____表示。
4. 检查长期计划执行情况时，如计划指标是按计划期末应达到的水平下达的，应采用_____计算；如计划指标是按全期累计完成量下达的，则采用_____计算。
5. 同类指标数值在不同空间进行静态对比的结果，就是_____，而同一总体内不同部分数值静态对比的结果，则是_____，二者是有区别的。
6. 把两个地区 2015 年粮食产量进行对比，这个相对指标是_____相对指标。
7. 2015 年深圳市全年旅游住宿设施接待过夜游客 4 991.06 万人次，这是_____总量指标。
8. 在经济生活中，表示食物支出金额占总支出金额百分比的恩格尔系数经常被作为衡量一个国家或地区贫困或富裕的重要指标，恩格尔系数属于_____相对指标。
9. 居民人均生活费收入与职工平均工资是两个不同的指标，前者是_____指标，后者是_____指标。
10. 在计算和应用相对指标时，要严格保持分子、分母的_____性。

五、应用能力训练题

1. 某公司在年终统计分析报告中写道：我公司今年销售收入计划规定 8 000 万元，实际完成了 8 160 万元，超额完成计划 2%；销售利润率计划规定 8%，实际为 12%，超额完成计划 4%；劳动生产率计划规定比去年提高 5%，实际比去年提高 5.5%，完成计划 110%；产品单位成本计划规定比去年下降 3%，实际比去年下降 2.5%，实际比计划多下降 0.5 个百分点。

请指出上述分析报告中哪些指标计算有错误？将其改正过来。

2. 某市 2014 年和 2015 年地区生产总值资料如下表所示，请计算并填写列表中所缺数字。

项 目	2015 年		2014 年实际完成（亿元）	2015 年比 2014 年增长（%）
	实际完成（亿元）	比重（%）		
地区生产总值	2 100			15
其中：第一产业		7.5	140	
第二产业			900	20
第三产业				

3. 下表中各指标是媒体中经常出现的指标，请指出它们的类别（请在对应栏目下划√）。

序号	指标名称	时期 指标	时点 指标	计划完 成程度	结构相 对指标	比例相 对指标	比较相 对指标	动态相 对指标	强度相 对指标
1	进出口总额								
2	在校学生数								
3	人均国内生产总值								
4	我国 2014 年全年国内游客 36.1 亿人次，比 上年增长 10.7%								
5	2014 年按实际汇率计中国 GDP 总量相当于 美国的 61.7%								
6	人口性别比例								
7	升学率								
8	实际销售量是计划销售量的 198%								
9	出生人口数								
10	农民人均纯收入								
11	2014 年按购买力平价计，中国 GDP 总量达 19.230 万亿美元，高于美国 0.943 万亿美元								
12	2014 年我国移动电话普及率 94.5 部/百人								
13	全市医保定点零售药店数								
14	商品房中销售给个人的占 93%								
15	2014 年全国普通高等教育本专科招生 721.40 万人，比上年增长 3.08%								
16	2014 年国家全员劳动生产率 72 313 元/人								
17	2014 年全国招收硕士研究生 62.13 万人								
18	恩格尔系数								
19	从业人员数								
20	每百人拥有的私人轿车数								

4. 某公司 2015 年计划产值为 3 160 万元，结果计划完成了 110%，2015 年产值计划比 2014 年增长 8.5%，试计算实际产值 2015 年比 2014 年的增长率。

5. 某公司产值计划完成 103%，比上年增长 6%，请问，计划规定比上年增加多少？又知该公司某款主要产品每台成本应在 2014 年 1 398 元的水平上降低 20 元，实际 2015 年每台成本为 1 365 元，试确定降低成本计划完成情况指标。

6. 某生产公司 2014 年经过战略调整，改革措施到位，公司取得了辉煌的业绩。在年终总结报告中，公司管理人员生动、丰富又有说服力地表达了生产业绩的“好”。下面是部分摘录内容。

2014 年我公司实现产量 110 万台，这是公司历史上的最高产量。与我们年初定的计划 90 万台相比，我们超额 22.2%完成了任务，多生产了 20 万台；与我们去年产量相比，今年我们再上一个新台阶，比去年年产 85 万台的产量多 29.4%，比去年多生产了 25 万台。这是一个令人骄傲的成绩，因为我们同行业的竞争对手 A 厂最高水平是 108 万台，我们已经超出 1.85%，还

多出2万台。这个成绩给了我们很强的信心，我们“十二五”计划总产量是400万台，仅2014年一年完成的产量就占五年计划的27.5%，累计到2014年年末，我们已经生产340万台，完成“十二五”计划的85%，提前完成五年计划已成定局。但我们在成绩面前要保持清醒，我们公司非生产工人人数较多，所以，全员劳动生产率指标还不算太高，我们应该奋发进取，不断创新，争取更大成绩！

写出上述文字中出现的所有指标，它们分别属于哪一种总量指标？哪一种相对指标？通过此例，拓展思路，学会在实际应用时纵横对比，多角度、全方位描述现象。



本章案例

2015年上半年，面对复杂的国内外经济环境和不断加大的下行压力，党中央、国务院坚持稳中求进的工作总基调，科学精准实施宏观调控，坚定不移推进体制改革与制度创新，国民经济运行处在合理区间，主要指标逐步回暖，呈现缓中趋稳、稳中有好的发展态势。

初步核算，上半年国内生产总值为296 868亿元，按可比价格计算，同比增长7.0%。分季度看，一季度同比增长7.0%，二季度增长7.0%。分产业看，第一产业增加值为20 255亿元，同比增长3.5%；第二产业增加值为129 648亿元，增长6.1%；第三产业增加值为146 965亿元，增长8.4%。从环比看，二季度国内生产总值增长1.7%。

一、农业生产形势较好

全国夏粮总产量为14 107万吨，比上年增加447万吨，增长3.3%。夏收油菜籽产量为1 388万吨，比上年增加16万吨，增长1.2%。上半年，猪、牛、羊、禽肉产量为3 906万吨，同比下降2.4%，其中猪肉产量为2 574万吨，下降4.9%。

二、工业生产基本平稳

上半年，全国规模以上工业增加值按可比价格计算同比增长6.3%，增速比一季度回落0.1个百分点。分经济类型看，国有控股企业增加值同比增长1.9%，集体企业增长2.0%，股份制企业增长7.5%，外商及港、澳、台商投资企业增长3.8%。分三大门类看，采矿业增加值同比增长3.2%，制造业增长7.1%，电力、热力、燃气及水生产和供应业增长2.2%。分产品看，565种产品中有305种产品产量同比增长。上半年规模以上工业企业产销率达到97.3%。规模以上工业企业实现出口交货值55 707亿元，同比下降0.4%。6月份，规模以上工业增加值同比增长6.8%，增速连续3个月回升，环比增长0.64%。

1~5月份，全国规模以上工业企业实现利润22 548亿元，同比下降0.8%。规模以上工业企业每百元主营业务收入中的成本为85.95元，主营业务收入利润率为5.38%。

三、固定资产投资增速有所回落

上半年，固定资产投资（不含农户）237 132亿元，同比名义增长11.4%（扣除价格因素实际增长12.5%），增速比一季度回落2.1个百分点。其中，国有控股投资73 745亿元，增长12.3%；民间投资154 438亿元，增长11.4%，占全部投资的比重为65.1%。分产业看，第一产业投资6 159亿元，同比增长27.8%；第二产业投资97 446亿元，增长9.3%；第三产业投资133 527亿元，增长12.4%。从到位资金情况看，上半年到位资金为261 507亿元，同比增长6.3%。其中，国

家预算资金增长 18.6%，国内贷款下降 4.8%，自筹资金增长 8.6%，利用外资下降 30.9%。上半年新开工项目计划总投资 191 936 亿元，同比增长 1.6%。从环比看，6 月份固定资产投资（不含农户）增长 0.88%。

上半年，全国房地产开发投资 43 955 亿元，同比名义增长 4.6%（扣除价格因素实际增长 5.7%），增速比一季度回落 3.9 个百分点，其中住宅投资增长 2.8%。房屋新开工面积 67 479 万平方米，同比下降 15.8%，其中住宅新开工面积下降 17.3%。全国商品房销售面积 50 264 万平方米，同比增长 3.9%，一季度为同比下降 9.2%，其中住宅销售面积增长 4.5%。全国商品房销售额为 34 259 亿元，同比增长 10.0%，其中住宅销售额增长 12.9%。房地产开发企业土地购置面积为 9 800 万平方米，同比下降 33.8%。6 月末，全国商品房待售面积为 65 738 万平方米，同比增长 20.8%。上半年，房地产开发企业到位资金为 58 948 亿元，同比增长 0.1%。

四、商品消费稳健增长

上半年，社会消费品零售总额为 141 577 亿元，同比名义增长 10.4%（扣除价格因素实际增长 10.5%），增速比一季度回落 0.2 个百分点。其中，限额以上单位消费品零售额为 66 256 亿元，增长 7.4%。按经营单位所在地分，城镇消费品零售额为 121 850 亿元，同比增长 10.2%，乡村消费品零售额为 19 727 亿元，增长 11.6%。按消费形态分，餐饮收入 14 996 亿元，同比增长 11.5%；商品零售 126 581 亿元，增长 10.3%，其中限额以上单位商品零售 62 306 亿元，增长 7.4%。6 月份，社会消费品零售总额同比名义增长 10.6%（扣除价格因素实际增长 10.6%），增速比 5 月份加快 0.5 个百分点，环比增长 0.96%。

上半年，全国网上零售额为 16 459 亿元，同比增长 39.1%。其中，实物商品网上零售额为 13 759 亿元，增长 38.6%，占社会消费品零售总额的比重为 9.7%；非实物商品网上零售额为 2 700 亿元，增长 41.9%。

五、对外贸易顺差继续增加

上半年，进出口总额为 115 316 亿元人民币，同比下降 6.9%。其中，出口 65 722 亿元人民币，增长 0.9%；进口 49 594 亿元人民币，下降 15.5%。进出口相抵，顺差 16 128 亿元人民币。6 月份，进出口总额为 20 655 亿元人民币，同比下降 1.9%。其中，出口 11 749 亿元人民币，增长 2.1%；进口 8 907 亿元人民币，下降 6.7%。

六、居民消费价格基本稳定

上半年，居民消费价格同比上涨 1.3%，涨幅比一季度扩大 0.1 个百分点。其中，城市上涨 1.3%，农村上涨 1.1%。分类别看，食品价格同比上涨 2.0%，烟酒及用品上涨 0.5%，衣着上涨 2.9%，家庭设备用品及维修服务上涨 1.1%，医疗保健和个人用品上涨 1.8%，交通和通信下降 1.6%，娱乐教育文化用品及服务上涨 1.5%，居住上涨 0.7%。在食品价格中，粮食价格上涨 2.6%，油脂价格下降 4.4%，猪肉价格上涨 2.4%，鲜菜价格上涨 4.6%。6 月份，居民消费价格同比上涨 1.4%，环比持平。上半年，工业生产者出厂价格同比下降 4.6%，6 月份同比下降 4.8%，环比下降 0.4%。上半年，工业生产者购进价格同比下降 5.5%，6 月份同比下降 5.6%，环比下降 0.2%。

七、居民收入稳定增长

根据城乡一体化住户调查，上半年全国居民人均可支配收入为 10 931 元，同比名义增长 9.0%，扣除价格因素实际增长 7.6%。按常住地分，城镇居民人均可支配收入为 15 699 元，同

比名义增长 8.1%，扣除价格因素实际增长 6.7%；农村居民人均可支配收入为 5 554 元，同比名义增长 9.5%，扣除价格因素实际增长 8.3%。全国居民人均可支配收入中位数为 9 700 元，同比名义增长 10.5%。二季度末，农村外出务工劳动力总量为 17 436 万人，同比增加 18 万人，增长 0.1%。上半年，外出务工劳动力月均收入 3 002 元，同比增长 9.8%。

八、结构继续优化

产业结构继续优化。上半年，第三产业增加值占国内生产总值的比重为 49.5%，比上年同期提高 2.1 个百分点，高于第二产业 5.8 个百分点。内需结构进一步改善。上半年，最终消费支出对国内生产总值增长的贡献率为 60.0%，比上年同期提高 5.7 个百分点。城乡居民收入差距进一步缩小。上半年，农村居民人均可支配收入实际增长快于城镇居民人均可支配收入 1.6 个百分点，城乡居民人均收入倍差 2.83，比上年同期缩小 0.04。节能降耗继续取得新进展。上半年，单位国内生产总值能耗同比下降 5.9%。

九、货币信贷平稳增长

6 月末，广义货币（M2）余额为 133.34 万亿元，同比增长 11.8%；狭义货币（M1）余额为 35.61 万亿元，增长 4.3%；流通中货币（M0）余额为 5.86 万亿元，增长 2.9%。6 月末，人民币贷款余额为 88.79 万亿元，人民币存款余额为 131.83 万亿元。上半年，新增人民币贷款 6.56 万亿元，同比多增 5 371 亿元；新增人民币存款 11.09 万亿元，同比少增 3 756 亿元。上半年，社会融资规模增量为 8.81 万亿元。

总之，在党中央、国务院一系列政策措施作用下，二季度主要指标增速企稳向好，国民经济出现积极变化，经济发展活力动力增强。但也要看到，国内外经济环境依然错综复杂，全球经济复苏曲折缓慢，我国经济回稳基础还需进一步巩固。下一阶段，要认真贯彻落实党中央、国务院各项决策部署，坚定信心，真抓实干，正确处理好稳增长、促改革、调结构、惠民生和防风险的关系，促进我国经济发展速稳质优，保持中高速，迈向中高端。

（摘自国家统计局中国统计信息网 2015-07-15）

平均指标和标志变异指标

学习要点

- 平均指标的意义和作用。
- 数值平均数的计算方法。
- 位置平均数的含义及适用范围。
- 运用标志变异指标进行平均数代表性的对比分析。

指标是反映总体特征的。总量指标反映总体、总规模、总水平和工作总量；相对指标反映总体各现象之间数量对比关系；而要反映总体在具体条件下最集中的趋势，则需要运用平均指标。

平均指标我们都不陌生，我们经常接触它，根据我们的理解去使用它。无论是在工作学习中，还是在日常生活中，有关平均指标的资料比比皆是。例如，2014 年全国城镇非私营单位就业人员年平均工资为 56 339 元，与 2013 年的 51 483 元相比，增加了 4 856 元，同比名义增长 9.4%，增幅回落 0.7 个百分点。其中，在岗职工年平均工资为 57 346 元，同比名义增长 9.5%，增幅回落 0.6 个百分点（资料来源：国家统计局网站 2015-05-27）。我们正是通过对非私营单位及在岗职工年平均工资的描述，获得关于全国城镇非私营单位及其在岗职工的年收入水平信息，并通过观察其增长速度，反映城镇职工群体生活状况。这种利用统计指标获得的反映总体的集中趋势（最具普遍意义的代表水平）的信息，能为我们正确认识问题、解决问题提供科学的依据。

5.1 平均指标的意义和种类

5.1.1 平均指标的意义和作用

平均指标又称平均数，反映现象总体各单位某一数量标志值的典型水平、一般水平和代表性水平。平均指标是社会经济现象中最常用的一种综合指标。

例如，我们经常计算的某门课的班级平均分数，就是全班同学成绩的一般水平。在全班同学总体中，每位同学是总体单位，成绩 80 分、90 分、78 分等都是数量标志值。大家成绩参差不齐，各不相同，如何反映全班同学总体水平？经过计算，用平均分数来反映。平均分数是全



班同学成绩的最一般水平、最典型水平，最具有代表性。

平均指标的显著特点是：它不是某一单位的具体数值，而是代表总体某种数量标志的一般水平，是总体各单位的代表值；平均指标把总体各单位标志值的差异给抽象化了，它是一个抽象化的数值。正是由于平均指标的“抽象化”特征，当我们计算出某地平均每户人口规模 2.94 人时，不必对数值进行四舍五入。

由于平均指标能够综合反映某种社会经济现象总体在一定条件下的一般水平，所以它应用很广泛，其作用主要表现在以下几个方面。

1. 可以反映总体各单位分布的集中趋势

在社会经济现象总体中，各单位中某一数量标志值往往各不相同，通常情况下，标志值很大或标志值很小的单位比较少，而逐渐靠近平均数的单位数逐渐增多，标志值围绕在平均数周围的单位占最大比重。平均指标反映了标志值变动的集中趋势。例如，2014 年上海职工月平均工资为 5 451 元就反映了职工工资水平的集中趋势，是职工工资在 2014 年所达到的一般水平。

2. 可以对现象在不同空间、时间上进行比较分析

对于不同国家、不同地区、不同单位的同类现象的水平，由于总体范围的大小可能不同，通常不能直接进行比较分析，然而通过计算平均指标，可以使现象具有可比性。例如，甲乙两个公司 12 月份职工工资总额分别是 300 万元和 630 万元，职工人数分别为 400 人和 1 000 人，尽管从总量指标“工资总额”和“职工人数”看，乙公司均高于甲公司，但从“人均工资”这一平均指标看，甲公司要高于乙公司 1 200 元/人。这个结果表明，甲公司工资水平高于乙公司。这个指标是通过平均指标的空间对比获得的。

根据同样道理，利用平均指标，可以对同一现象在不同时间进行比较分析。例如，将现在职工平均工资与改革开放前的工资水平进行比较分析，从中显示出职工工资水平的大幅度提高。

3. 可以分析现象之间的依存关系

由于平均指标是代表总体某种数量标志的一般水平，在对现象总体进行分组的基础上，可以通过计算各组的平均指标分析现象之间的依存关系。

4. 可以作为评价事物的参考依据

事物的好坏是相对的，通常以代表事物发展一般水平的平均指标为评定依据。例如，考生成绩为 78 分，如何判断成绩的高低？若平均成绩为 60 分，则 78 分比较高；若平均成绩为 90 分，则 78 分比较低。

5. 可以进行数量上的估算

在统计估计和预测中，往往利用并根据部分单位标志值计算的平均指标推算总体平均指标，或以总平均指标推算总体标志总量。例如，已知市内某区牛奶的人均消费量，可以采用科学方法，推断全市的人均牛奶消费量，也可以推断全市牛奶的消费总量。

5.1.2 平均指标的种类

平均指标按其所属总体的时间范围不同分为两种：静态平均数和动态平均数。静态平均数是反映同一时间范围内总体各单位某一数量标志一般水平的平均指标；动态平均数是反映不同时间而同一空间范围内总体某一数量标志一般水平的平均指标。本章主要讨论静态平均数，动态平均数将在第7章的时间数列中专门讨论。

静态平均数按其计算方法的不同分为两种：数值平均数和位置平均数。

根据总体各单位标志值计算的平均数，称为数值平均数，常见的主要有算术平均数、调和平均数和几何平均数等。根据总体标志值在分配数列中的位置确定的平均数，称为位置平均数，常见的主要有众数、中位数和四分位数等。

5.2 数值平均数

数值平均数按照计算方法不同，分为算术平均数、调和平均数和几何平均数。

5.2.1 算术平均数

1. 算术平均数的意义

算术平均数（arithmetic mean）是总体标志总量与总体单位总量对比的结果，其基本公式为：

$$\text{算术平均数} = \frac{\text{总体标志总量}}{\text{总体单位总量}} \quad (5-1)$$

算术平均数与强度相对指标都是比值，都有“平均”的含义，但两者的明显区别在于：算术平均数的分子和分母是同一个总体的两个总量指标，分子是标志总量，分母是单位总量，而且分子、分母位置不能互换；强度相对指标的分子和分母分属两个不同总体的总量指标，而且分子、分母位置颠倒有意义，它有正、逆指标之分。例如，一个三口之家，父亲和母亲均在岗，每月两人工资总额为12 000元，孩子上小学。则12 000元除以2，表示夫妻平均工资6 000元/人；12 000元除以3，表示全家人均收入4 000元/人。在这里，孩子没有收入，但分母单位总量里包含他，则分母总体扩大，与分子是不同的两总体。所以，“人均收入”是强度相对指标，而“平均工资”才是同一总体的标志总量与单位总量之比，是平均指标。

算术平均数是分析社会经济现象一般水平和典型特征的最基本、最常用的一种平均指标，它也是平均指标中最重要的一种。一般不加特别说明时，所称的“平均数”都是指算术平均数。

根据基本公式计算算术平均数时，由于依据的资料不同，计算方法有所不同，可分为简单算术平均数和加权算术平均数。

2. 简单算术平均数

如果总体单位数不多，资料未经分组，计算算术平均数时可采用简单算术平均数的方法。

计算过程：先把总体各单位标志值相加得总体标志总量，再与总体单位总数对比，即可得到算术平均数。计算公式为：

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum x}{n} \quad (5-2)$$

式中： \bar{x} ——算术平均数；

x ——总体各单位标志值；

n ——总体单位数；

\sum ——加总的符号。

【实例 5.1】 一个公司有 5 个部门，每个部门员工数分别为 24、13、19、26 和 11，求平均每部门的人数。

解：

$$\text{平均人数} = \frac{24+13+19+26+11}{5} = \frac{93}{5} = 18.6 \text{ (人)}$$

3. 加权算术平均数

如果调查所得的原始资料已经经过分组整理，形成了变量数列，则计算算术平均数要采用加权算术平均数的方法。计算过程：将各组的变量值与各组的单位数相乘，计算出各组标志总量，将各组标志总量汇总得出总体标志总量，然后除以各组单位数之和，即总体单位总量，得到平均数。计算公式为：

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \cdots + x_n f_n}{f_1 + f_2 + f_3 + \cdots + f_n} = \frac{\sum xf}{\sum f} \quad (5-3)$$

式中： x ——各组标志值；

f ——各组次数。

【实例 5.2】 服装商店要销售 100 件毛衣，其中 20 件大号毛衣，每件 200 元，50 件中号毛衣，每件 190 元，30 件小号毛衣，每件 180 元。计算每件毛衣平均价格。

解：根据题意，可列出如表 5.1 所示的计算表。

表 5.1 服装商店毛衣销售价格资料及其计算表

销售价格 x (元)	件数 f (件)	销售总价值 xf (元)
200	20	4 000
190	50	9 500
180	30	5 400
合 计	100	18 900

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{18\,900}{100} = 189 \text{ (元)}$$

由上式可见，平均数的大小不仅取决于总体各单位标志值 x ，同时也取决于各标志值的次数 f 。次数多的标志值对平均数的影响要大些，次数少的标志值对平均数的影响相应就小些。标志值次数的多少，对平均值的大小有权衡轻重的影响作用，所以称其为权数。

权数不但可以用次数、频数（即总体各组单位数）这种绝对数表示，还可以用比重、频率这种相对数表示。此时，加权算术平均数公式可以演化为：

$$\bar{x} = \sum x \frac{f}{\sum f} \tag{5-4}$$

仍以表 5.1 资料为例，各型号毛衣价格及件数构成资料如表 5.2 所示。

表 5.2 各型号毛衣价格及件数构成资料

销售价格 x (元)	件数构成 $\frac{f}{\sum f}$ (%)	$x \frac{f}{\sum f}$
200	20	40
190	50	95
180	30	54
合 计	100	189

解：

$$\bar{x} = \sum x \frac{f}{\sum f} = 189 \text{ (元)}$$

由上例明显看出，比重的大小直接表明了该组标志值占据平均数的地位。所以，权数对于算术平均数的影响作用，不是决定于权数本身数值的大小，而是决定于权数比重（或称为相对权数）的大小。权数比重是指作为权数的各组单位数占总体单位数的比重。哪一组的单位数所占的比重大，哪一组标志值对平均数的影响就大。

当各组的单位数相等或各组单位数所占的比重相等时，权数对各组的作用都一样，加权也就失去了意义。此时，加权算术平均数等于简单算术平均数。即当 $f_1=f_2=f_3=\cdots=f_n$ 时， $\bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum x}{n}$ ，可见，简单算术平均数是加权算术平均数的特例。

在计算加权算术平均数时，还会遇到权数的选择问题。选择权数的原则是，务必使各组的标志值与其乘积等于各组的标志总量，并且具有实际的经济意义。在分配数列条件下，一般来说，次数就是权数。但也有例外，特别是用相对数或平均数计算加权算术平均数时，要特别注意。

【实例 5.3】 某居住小区 60 户居民人均居住面积资料如表 5.3 所示，求总人均居住面积。

表 5.3 小区居民人均居住面积资料及其计算表

人均居住面积 x (平方米/人)	户 数	居住人数 f (人)	居住面积 xf (平方米)
60	3	6	360
40	47	141	5 640
30	4	16	480
24	6	30	720
合 计	60	193	7 200

本例要平均的对象是各户的人均居住面积，为计算小区总人均居住面积，应该以居住人数为权数，户数虽是不同人均居住面积的次数，但并不是合适的权数。因为人均居住面积与居住人数相乘才有具体的经济含义，即为居住总面积。计算结果如下：

解：

$$\text{小区总人均居住面积 } \bar{x} = \frac{\sum xf}{\sum f} = \frac{7\,200}{193} = 37.31 \text{ (平方米/人)}$$

上面的例题采用的都是单项式数列，如果是组距式分配数列，则先计算组中值来代表各组的标志值 x ，再计算平均数。但应用这种计算方法需有一个假定条件，即假定各单位标志值在各组内是均匀分布的，而实际上，这是很难达到的。所以，这种用组中值来代替计算的算术平均数，不可避免地会存在一定程度的误差，具有近似值的性质。

【实例 5.4】 某班学生“统计学”测验成绩被编制成组距式分配数列，如表 5.4 所示，计算班级平均成绩。

表 5.4 “统计学”测验成绩资料及其计算表

成绩 (分)	学生数 f	组中值 x	xf
60 以下	2	55	110
60~70	12	65	780
70~80	21	75	1 575
80~90	10	85	850
90 及以上	5	95	475
合 计	50	—	3 790

解：

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{3\,790}{50} = 75.8 \text{ (分)}$$

5.2.2 调和平均数

调和平均数 (harmonic mean) 是各个标志值倒数的算术平均数的倒数，故又称为倒数平均数。根据给定的材料不同，调和平均数有简单调和平均数和加权调和平均数两种。

1. 简单调和平均数

简单调和平均数是各个标志值倒数的简单算术平均数的倒数。其计算公式为：

$$\overline{x_h} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x}} \quad (5-5)$$

式中： $\overline{x_h}$ ——调和平均数；

x ——各标志值；

n ——项数。

【实例 5.5】 3 个蔬菜超市销售同一种蔬菜，但价格不同，每 500 克蔬菜价格分别为 1.8 元、2 元、2.3 元。若在 3 个超市各购买 1 元钱的这种蔬菜，则蔬菜的平均价格为多少？

解：

$$\text{平均价格 } \overline{x_h} = \frac{n}{\sum \frac{1}{x}} = \frac{1+1+1}{\frac{1}{1.8} + \frac{1}{2} + \frac{1}{2.3}} = 2.012 \text{ (元)}$$

上式中，蔬菜平均价格的计算采用了新的计算方法，因为给出的资料与计算算术平均数的资料不同。在这里，有各组的标志值（即价格），但没有各组的次数（购买量），却有两者的乘积（购买额），所以分母购买量利用购买额除以价格得到。可见，调和平均数与算术平均数计算原理相同，有观点说，调和平均数是算术平均数的变形。

2. 加权调和平均数

加权调和平均数是各个标志值倒数的加权算术平均数的倒数。在实例 5.5 中，各组购买额恰好相等，均为 1 元钱，如果购买额不相等，则必须采用加权调和平均数，其计算公式为：

$$\overline{x_h} = \frac{m_1 + m_2 + m_3 + \cdots + m_n}{\frac{m_1}{x_1} + \frac{m_2}{x_2} + \frac{m_3}{x_3} + \cdots + \frac{m_n}{x_n}} = \frac{\sum m}{\sum \frac{m}{x}} \quad (5-6)$$

式中： m ——总体各组标志总量。

【实例 5.6】 学校食堂购进某种蔬菜，相关资料如表 5.5 所示，求蔬菜的平均价格。

表 5.5 蔬菜价格资料及其计算表

购买地点	价格 x (元/500 克)	购买金额 m (元)	购买量 m/x (500 克)
甲超市	1.8	20	11.11
乙超市	2	15	7.5
丙超市	2.3	10	4.35
合 计	—	45	22.96

解：

$$\text{蔬菜的平均价格 } \bar{x}_h = \frac{\sum \frac{m}{x}}{\sum \frac{m}{x}} = \frac{45}{22.96} = 1.96 \text{ (元)}$$

调和平均数与算术平均数在计算上是相通的，但各自适合不同的资料。在实际应用中，一般采用下列步骤进行计算。

(1) 在给出的材料中找出 x 。方法是“求谁的平均数谁就是 x ”。

(2) 确定权数为 f 还是为 m 。方法是分配数列中的这栏数值若与 x 相乘有意义，则为 f ；若与 x 相除有意义，则为 m 。

(3) 写出计算公式。若权数定为 f ，公式为

$$\bar{x} = \frac{\sum xf}{\sum f}$$

即采用算术平均数；若权数定为 m ，则公式为

$$\bar{x} = \frac{\sum \frac{m}{x}}{\sum \frac{m}{x}}$$

即采用调和平均数。

(4) 延展计算表。根据写出的公式，需要什么内容就在表格中列出什么。例如，若采用的是算术平均数公式，则表格中列出 xf 以便于计算；若采用的是调和平均数公式，则表格中列出 $\frac{m}{x}$ 以便于计算。

(5) 将数值代入公式，进行计算，得到答案。

5.2.3 几何平均数

几何平均数 (geometric mean) 是若干项变量值的连乘积开若干次项数的方根。它是计算平均数的另一种形式，主要用于计算比率或速度的平均。当所掌握的变量值本身是比率的形式，而且各比率的乘积等于总的比率时，就采用几何平均法计算平均比率。因为这与几何平均数的数学性质相一致。根据所掌握的资料不同，几何平均数分为简单几何平均数和加权几何平均数。

1. 简单几何平均数

如果资料未分组，直接将 n 项变量值连乘，然后对其连乘积开 n 次方所得的平均数即为简单几何平均数。其计算公式为：

$$\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n} = \sqrt[n]{\prod x} \quad (5-7)$$

式中： \bar{x}_G ——几何平均数；

x ——各标志值;

n ——变量值个数;

\prod ——连乘符号。

【实例 5.7】 产品的生产往往需要几道生产工序, 只有在第一道工序合格的产品才能进入第二道工序。现已知纺织厂纺纱车间产品合格率为 91%, 织布车间产品合格率为 89%, 印染车间产品合格率为 87%, 求 3 个车间平均产品合格率。

解: 由于后续车间的合格率是在前一车间产品合格基础上计算的, 所以全厂产品的总合格率并不等于各车间产品合格率的总和, 而是各车间产品合格率的连乘积, 因此要采用几何平均法计算各车间产品平均合格率, 即产品平均合格率为:

$$\overline{x_G} = \sqrt[n]{x_1 \square x_2 \square x_3 \square \cdots \square x_n} = \sqrt[3]{91\% \times 89\% \times 87\%} = 88.99\%$$

2. 加权几何平均数

与算术平均数一样, 当资料已经经过整理, 应以各变量值出现的次数为权数, 计算加权几何平均数。其计算公式为:

$$\overline{x_G} = \sqrt[\sum f]{x_1^{f_1} \square x_2^{f_2} \square x_3^{f_3} \square \cdots \square x_n^{f_n}} \quad (5-8)$$

式中: f ——各变量值次数;

$\sum f$ ——次数总和。

【实例 5.8】 一条产品流水线由 12 道工序组成, 其中, 合格率为 98% 的有 2 道工序, 合格率为 96% 的有 5 道工序, 合格率为 92% 的有 3 道工序, 合格率为 89% 的有 2 道工序。求产品总平均合格率。

解: 产品总平均合格率为:

$$\overline{x_G} = \sqrt[12]{98\%^2 \times 96\%^5 \times 92\%^3 \times 89\%^2} = 94.12\%$$

5.3 位置平均数

数值平均数都是根据标志值计算得到的, 而位置平均数是根据总体各单位标志值所处的位置确定的。位置平均数主要包括众数、中位数和四分位数。

5.3.1 众数

1. 众数的意义

众数(mode)是指总体中出现次数最多的标志值, 是总体各单位一般水平的代表值, 反映现象的集中趋势, 用 M_o 表示。众数可能不存在或不唯一。

在实际工作中，众数的应用比较广泛。例如，为掌握消费者需要的鞋的号码大小，不需要全面登记所有鞋码进行平均，只用生活中最普遍的、成交量最大的尺码，它反映了人们一般的需求。

2. 众数的确定

众数的确定方法根据具体资料而定。

（1）由未分组资料确定众数。在资料未分组的情况下，众数的确定很简单，只需找出次数最多的标志值即可。例如，一组学生年龄分别为 18、19、19、20、20、20、20、22，则众数为 20；若学生年龄为 18、19、19、19、20、20、20、22，则有双众数，分别是 19、20；若学生年龄为 16、17、18、19、20、21、22、23，则不存在众数。

（2）由单项式数列确定众数。在单项数列情况下，次数最多的组的标志值便是众数，如表 5.6 所示。

表 5.6 某商场销售成年女鞋资料表

女鞋号码	销售量（百双）
35 码	1.1
36 码	2
37 码	5
38 码	2.1
39 码	0.6
40 码	0.2
合 计	10.9

从表 5.6 中可以看出，出现次数最多（5 百双）的变量值是 37 码的鞋，因此，37 码就是该商场女鞋销售的众数。

（3）由组距数列确定众数。在组距数列情况下，计算众数的方法是：首先在数列中找到次数最多的组；然后，用公式计算众数的近似值，计算公式如下。

$$\text{下限公式: } M_o = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times d \quad (5-9)$$

$$\text{上限公式: } M_o = U - \frac{\Delta_2}{\Delta_1 + \Delta_2} \times d \quad (5-10)$$

式中：M_o——众数；

L——众数所在组的下限；

U——众数所在组的上限；

Δ₁——众数所在组次数与前一组次数之差；

Δ₂——众数所在组次数与后一组次数之差；

d——众数所在组的组距。

【实例 5.9】 某班 30 名大学生的体重资料如表 5.7 所示，求 30 名大学生的体重的众数。

表 5.7 某班 30 名大学生的体重资料

体重（千克）	人数（人）
44~48	2
48~52	3
52~56	6
56~60	10
60~64	7
64~68	2
合 计	30

解：为求众数，首先确定数列的众数组。例 5.9 中次数最多的组是“56~60”组，按照下限公式，众数值近似确定为：

$$M_0 = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times d = 56 + \frac{4}{4 + 3} \times 4 = 58.29 \text{（千克）}$$

若将资料带入上限公式，计算结果与下限公式结果一样。

3. 众数的特点及应用

（1）众数是根据变量值出现的次数确定的，而不是通过所有变量值计算得到的，所以，众数不受到极端变量值的影响。众数的这一特点是数值平均数所不具备的。在实际工作中，众数用得最多的是具有明显偏态集中趋势的次数分配。例如，按照统计国际惯例，对家庭收入分配数列、工人周工资分配数列、某种债券息票率分组的行情次数等进行的分析，都采用出现次数最多的众数，得到“最普通的家庭收入额”，“最普通的工人周工资金额”，“最常见的外汇率、息票率”等。

（2）众数是出现次数最多的变量值，如果数据分布没有明显集中趋势，众数就可能不存在；如果有两个最高次数，也可以有两个众数（bimodal）。只有在总体单位比较多，而且又明显地集中于某个变量值时，众数计算才有意义。

（3）众数主要用于测度定类数据的集中趋势，当然也适用于作为定序数据及定距和定比数据集中趋势的测度值。众数是唯一一个能用在名词数据上的平均数。例如，一项对大学生的研究包括了 20 个心理学专业的学生、80 个英语专业的学生和 16 个管理专业的学生。我们无法计算这些专业的平均数，但我们可以指出众数是英语专业，因为它是出现次数最多的专业。在这里，“英语专业”是定类数据的集中趋势。

5.3.2 中位数和四分位数

1. 中位数的意义

中位数（median）是指将总体各单位标志值按一定顺序排列后，处于中间位置的标志值。

由于它的位置居中，其数值不受极端数值的影响，也能表明总体标志值的一般水平，用 M_e 表示。中位数将全部标志值分成两半，一半小于中位数，一半大于中位数，所以中位数又称为二分位数。

在实际工作中，有许多场合用中位数来表示现象的一般水平。例如，在研究居民收入水平时，以居民收入中位数来代表居民收入水平比采用算术平均数进行计算更为科学。

2. 中位数的确定

中位数的确定方法根据具体资料而定。

(1) 由未分组资料确定中位数。在资料未分组的情况下，中位数的确定比较简单。首先，把标志值按大小顺序排列起来，然后计算中位数所在的位置，找出中位数。如果总体单位数是奇数，处于 $\frac{n+1}{2}$ （ n 代表总体单位数）位置的标志值就是中位数；如果总体单位数是偶数，那么中位数就是处于 $\frac{n}{2}$ 与 $\frac{n}{2}+1$ 位置的两个标志值的算术平均数。

例如，某班甲、乙两个小组各有 7 人和 8 人，学生身高资料如下：

甲组学生身高（cm）——154、160、162、164、165、166、168；

乙组学生身高（cm）——157、159、162、163、164、167、168、172。

甲组中，中位数位置是 $\frac{7+1}{2}=4$ ，中位数是 164cm。

乙组中，中位数位置是 4（ $\frac{8}{2}=4$ ）与 5（ $\frac{8}{2}+1=5$ ）之间，中位数是 $\frac{163+164}{2}=163.5$ cm。

(2) 由单项式数列确定中位数。由单项式数列确定中位数时，首先将各组次数累计，然后根据公式 $\frac{\sum f+1}{2}$ 确定中位数的位次，最后从次数累计中找到中位数所在的组，该组的标志值就是中位数。

【实例 5.10】 某学院电子商务专业两个班学生参加某项技能测试，在规定的时间内完成任务数各不相同，资料如表 5.8 所示，求每班学生完成任务个数的中位数。

表 5.8 电子商务专业两个班学生某项技能测试完成任务个数资料

按完成任务数分组	一班学生数（人）	二班学生数（人）	一班人数累计（人）	二班人数累计（人）
1	3	4	3	4
2	5	5	8	9
3	8	6	16	15
4	7	8	23	23
5	5	7	28	30
合 计	28	30	—	—

解：一班学生完成任务数的中位数位次是 $\frac{\sum f+1}{2} = \frac{28+1}{2} = 14.5$ ，说明中位数位于第 14 个和第 15 个学生之间。由累计次数知，第 14 个和第 15 个学生都在第三组，标志值均为 3，所以，一班学生完成任务个数的中位数是 3。

二班学生完成任务数的中位数位次是 $\frac{\sum f+1}{2} = \frac{30+1}{2} = 15.5$ ，说明中位数位于第 15 个和第 16 个学生之间。由累计次数知，第 15 个和第 16 个学生分属第三组和第四组，所以应取这两组标志值的算术平均数作为中位数。所以，二班学生完成任务个数的中位数是 $\frac{3+4}{2} = 3.5$ 。

(3) 由组距数列确定中位数。在组距数列情况下，计算中位数的方法是：首先根据累计次数 $\frac{\sum f+1}{2}$ 确定中位数所在组，这个组的上下限就规定了中位数可能的取值范围；然后，用公式计算中位数的近似值，计算公式如下。

$$\text{下限公式: } M_e = L + \frac{\frac{\sum f}{2} - S_{m-1}}{f_m} \times d \quad (5-11)$$

$$\text{上限公式: } M_e = U - \frac{\frac{\sum f}{2} - S_{m+1}}{f_m} \times d \quad (5-12)$$

式中： M_e ——中位数；

L ——中位数所在组的下限；

U ——中位数所在组的上限；

f_m ——中位数所在组的次数；

S_{m-1} ——中位数所在组以前各组的累计次数；

S_{m+1} ——中位数所在组以后各组的累计次数；

$\sum f$ ——总次数；

d ——中位数所在组的组距。

【实例 5.11】 根据表 5.7 给出的某班 30 名大学生的体重资料，求大学生体重的中位数。

解：为求中位数，首先应求出中位数所在组，在总人数为 30 的前提下，中位数所在组为“56~60 千克”组。按照下限公式，中位数值近似确定为：

$$M_e = L + \frac{\frac{\sum f}{2} - S_{m-1}}{f_m} \times d = 56 + \frac{\frac{30}{2} - 11}{10} \times 4 = 57.6 \text{ (千克)}$$

若将资料带入上限公式，计算结果与下限公式结果一样。

3. 中位数的特点及应用

(1) 中位数是一种位置平均数,其大小取决于它在序列中的位置,因此它不受极端数值的影响。当存在极端数值时,中位数能比数值平均数更好地代表数据分布的一般水平。

(2) 中位数处于中间位置,有一半数值小于中位数,另一半数值大于中位数,所以,它能表明数字资料的集中趋势。在实际工作中,中位数用得较多的是测定人口年龄分配的平均年龄数。按照统计国际惯例,各国政府统计工作对年龄分组采用中位数,而不用算术平均数。因为人口的年龄分配不是中间高、两边低的分配形态,而是 J 形的分配形态。婴儿 0~1 岁的人数最多,随着年龄增大,人数逐渐减少,到百岁左右,所剩人数很少。若计算算术平均数,则会由于老人岁数很大的影响,使得算术平均年龄偏大,从而与实际情况不符。

(3) 中位数主要用于测度定序数据的集中趋势,当然也适用于定距数距和定比数据的集中趋势,但不适用于定类数据。

4. 四分位数

中位数是从中间点将全部数据等分为两部分,也称二分位数。同样,可通过三个数值,将全部标志值分割成为四个相等部分,其中每部分包含 25% 的数据,处在分位点上的这三个分割的数值就是四分位数 (quartiles),分别记为 Q_1 、 Q_2 和 Q_3 。很显然,中间的四分位数就是中位数,因此通常所说的四分位数是指第一个四分位数 Q_1 和第三个四分位数 Q_3 。

四分位数的确定方法与中位数类似,也要分为两种情况,即未分组资料和分组资料。对于未分组资料,首先将数据按大小顺序排序,然后计算分位数所在的位置,最后确定相关的四分位数。

【实例 5.12】 某公司 11 位员工六月份生产某种产品的数量分别是 17、16.8、16.6、16.9、18.4、17.7、18、18.2、17.5、16、18.6 公斤,计算四分位数。

解:将数据按大小顺序排列为:

16 16.6 16.8 16.9 17 17.5 17.7 18 18.2 18.4 18.6

分位数所在的位置分别是: Q_1 的位置 $= \frac{n+1}{4} = \frac{11+1}{4} = 3$;

$$Q_2 \text{ 的位置} = \frac{2(n+1)}{4} = \frac{n+1}{2} = \frac{11+1}{2} = 6;$$

$$Q_3 \text{ 的位置} = \frac{3(n+1)}{4} = \frac{3 \times (11+1)}{4} = 9。$$

因此,四分位数分别为 $Q_1=16.8$ (公斤), $Q_2=17.5$ (公斤), $Q_3=18.2$ (公斤)。

【实例 5.12】 中的资料项数 n 加上 1 后恰好为 4 的倍数,所以,四分位数的确定很简单。若 $n+1$ 不是 4 的整数倍数,按上述公式确定的四分位数的位置就带有小数,这时有关四分位数就应该是与该小数相邻的两个整数位置上的标志值的加权算术平均数,权数的大小取决于两个整数位置距离的远近。距离越近,权数越大;距离越远,权数越小,权数之和等于 1。

【实例 5.13】 某公司 10 位员工六月份生产某种产品的数量分别是 17、16.8、16.6、16.9、18.4、17.7、18、18.2、17.5、16 公斤，计算四分位数。

解：将数据按大小顺序排列为：

16 16.6 16.8 16.9 17 17.5 17.7 18 18.2 18.4

四分位数所在的位置分别是： Q_1 的位置 $= \frac{n+1}{4} = \frac{10+1}{4} = 2.75$ ；

$$Q_2 \text{ 的位置} = \frac{2(n+1)}{4} = \frac{n+1}{2} = \frac{10+1}{2} = 5.5；$$

$$Q_3 \text{ 的位置} = \frac{3(n+1)}{4} = \frac{3 \times (10+1)}{4} = 8.25。$$

则四分位数分别为：

$$Q_1 = 0.25 \times \text{第 2 项} + 0.75 \times \text{第 3 项} = 0.25 \times 16.6 + 0.75 \times 16.8 = 16.75 \text{ (千克)}；$$

$$Q_2 = 0.5 \times \text{第 5 项} + 0.5 \times \text{第 6 项} = 0.5 \times 17 + 0.5 \times 17.5 = 17.25 \text{ (千克)}；$$

$$Q_3 = 0.75 \times \text{第 8 项} + 0.25 \times \text{第 9 项} = 0.75 \times 18 + 0.25 \times 18.2 = 18.05 \text{ (千克)}。$$

由组距数列确定四分位数，必须先向上或向下累计次数，根据累计次数确定四分位数的位置，最后确定相关的四分位数。

四分位数不仅利用了中位数的基本方法，而且更加清晰地反映了被研究总体分布的集中情况。所以，四分位数也是描述集中趋势的重要指标。

与四分位数类似，还可以推算十分位数、百分位数。

在实际工作中，国家劳动部门采用“四分位法”测算工资指导线水平，即取下四分位值与上四分位值之间的幅度作为工资合理增长区间。政府运用工资指导线，对国有企业及其他各类企业的工资分配进行指导与调控，使企业工资增长符合经济和社会发展的要求，进一步促进生产力的发展。

5.3.3 应用平均指标要注意的问题

平均指标反映社会经济现象的集中趋势，在实际应用中，应注意以下几个问题。

1. 平均指标必须应用于同质总体

所谓同质总体，指总体具有同质性，即社会经济现象的各个单位在被平均的标志上具有同类性质。如果各单位在类型上存在根本差别，则依此计算的平均数不仅不能反映事物的本质和内在规律性，还会歪曲事物真相。例如，在研究职工工资收入变化时，如果把外来务工人员收入，或者是把企业高层管理者的收入与职工的工资收入合在一起求“平均”，则平均结果不能反映职工真实的工资收入变化。只有在同质总体基础上计算应用平均指标，才能有真实的社会经济意义。

2. 平均指标应与分配数列相结合

平均指标的重要特征是把总体各单位数量差异抽象化了，掩盖了各单位数量差别及其分布情况。所以在使用平均指标了解现象时，不能只看平均数，而要同时看分配数列的次数分布情况。所以，平均指标要与分配数列结合应用。

【实例 5.14】 某旅游公司对员工进行旅游英语知识测验，平均分数为 72.3 分，37 名员工成绩具体情况如表 5.9 所示。

表 5.9 旅游公司员工旅游英语测验成绩

旅游英语测验成绩（分）	员工人数（人）
60 以下	10
60~70	5
70~80	12
80~90	5
90~100	5
合 计	37

若单看平均数，只泛泛地了解了该旅游公司平均成绩 70 多分，比较不错。但如果结合分配数列，就会更具体、更深入地反映问题，即还有 10 人成绩不及格，占 37 名员工总数的 27%。

3. 平均指标要与分组法相结合

平均指标反映了总体各单位某一数量标志值的一般水平，但却掩盖了各组之间的差异。总体各组之间及各组内的差异往往影响总体的特征和分布规律，各组结构变动也会对总体变动产生影响。为了全面认识总体的特征和分布规律，需要将平均指标与统计分组结合起来。

【实例 5.15】 两种商品销售价格资料如表 5.10 所示，分别计算 6、7 月份两款家电平均销售价格并进行分析。

表 5.10 两种商品销售价格资料

商品名称	6 月销售价格（元）	6 月销售量（台）	7 月销售价格（元）	7 月销售量（台）
家电 I	100	60	90	40
家电 II	200	40	190	60

解：利用加权算术平均法计算，结果为：

6 月份两款家电平均销售价格= $\frac{100 \times 60 + 200 \times 40}{60 + 40} = 140$ （元）

7 月份两款家电平均销售价格= $\frac{90 \times 40 + 190 \times 60}{40 + 60} = 150$ （元）

7 月份两款家电平均销售价格比 6 月份两款家电平均销售价格上升了 10 元。但实际上，从按价格分组的资料看，无论是家电 I 还是家电 II，7 月份销售价格均比 6 月份低，即实际结果出现了与总平均数相矛盾的结论。之所以出现这种矛盾的结果，是因为两个月份相比，销售量的

结构发生了变化。7月份,价格较高的家电Ⅱ销售量所占比重为60%,比6月份时高出20%。正是由于这种权重结构的变化,导致了总平均数与实际情况相矛盾的情况出现。因此,在对实际中经济现象进行描述分析时,应将平均指标与分组法结合应用。

4. 选择合适的平均指标

在客观经济生活中,应该结合计算平均数的目的及被平均对象的特点选择正确恰当的平均指标。例如,民间的一则顺口溜就形象地反映了此问题:“张村有个张千万,隔壁九个穷光蛋,平均下来算一算,人人都是张百万。”此时应该选择众数或中位数来表示平均,而不是算术平均数。

5.4 标志变异指标

平均指标,包括数值平均数和位置平均数,它们都是现象一般水平的代表值,反映现象分布的集中趋势。但是,仅仅用集中趋势指标来描述现象的特征是不够的。

【实例 5.16】 在一次知识竞赛中,男、女两参赛代表队成绩资料如下。

男代表队: 51 65 69 75 81 87 94 95 96 97

女代表队: 74 76 78 79 82 82 83 84 86 86

这两参赛代表队的平均分数都是81分,但两队成绩的分散程度却不同。女代表队的成绩比较集中、整齐,即变异较小,从而平均数81分的代表性较好;男代表队的成绩比较分散,参差不齐,变动较大,用平均分81分作为代表,其代表性较差。统计平均指标在反映总体分布一般水平的同时,也掩盖了各标志值的差异性,而标志变异指标恰好弥补了这方面的不足。它能够综合反映总体分布中各标志值的差异程度,显示平均指标的代表性好坏,反映现象分布的离散趋势。

5.4.1 标志变异指标的意义和作用

1. 标志变异指标的意义

标志变异指标是反映总体各单位标志值差异程度的综合指标,又称为标志变动度。它是和统计平均指标相联系的一种综合指标。

平均指标将总体各单位标志值的差异抽象化,从一个侧面反映总体各单位标志值的集中趋势和程度。标志变异指标则从另一个侧面反映总体各单位标志值的差别大小、变动范围和离散程度。平均指标与标志变异指标分别反映同一总体在数量上的集中趋势与离散趋势,两者相辅相成,有助于科学、全面地描述客观现象特征,反映现象总体的数量规律。

2. 标志变异指标的作用

(1) 标志变异指标能够反映平均数代表性的大小。平均指标作为总体各单位标志值一般水

平的代表性指标，其代表性的大小与标志变异指标大小成反比。即标志变异指标越大，平均数的代表性越小；标志变异指标越小，平均数的代表性越大。

（2）标志变异指标可以说明现象变动的均衡性、稳定性和节奏性。标志变异指标越大，说明标志值之间的差异程度大，反映为总体均衡性、稳定性差，节奏感不强；标志变异指标越小，说明标志值之间的差异程度小，反映为总体均衡性、稳定性好，节奏感强。例如，企业为进行产品质量控制和实施全面质量管理，在考核生产计划执行情况时，除计算平均完成程度外，还研究计划执行的均衡性和节奏性。已知某公司所属两个企业 2015 年分季度计划完成情况资料如表 5.11 所示。

表 5.11 某公司所属企业 2015 年分季度计划完成情况资料表

公司所属企业	每季度完成全年计划程度（%）				合 计（%）
	一季度	二季度	三季度	四季度	
甲企业	24	25	25	26	100
乙企业	20	15	25	40	100

虽然两企业均完成全年计划任务，但甲企业各季度完成全年计划程度数值变动不大，说明其标志变动度小，则甲企业工作的均衡性和节奏性好；而乙企业各季度完成全年计划程度数值变动大，高为 40%，低仅有 15%，显示其标志变异指标大，所以，乙企业工作的均衡性和节奏性差。又如，在商品质量检测中，常用标志变异指标说明商品质量的稳定性。

（3）标志变异指标可以用来确定统计推断的准确程度。在抽样调查中，利用样本指标来推断总体指标时，标志变异指标数值的大小是确定统计推断准确程度及计算误差大小不可缺少的重要资料，具体内容将在第 6 章中详述。

5.4.2 标志变异指标的计算及应用

常用的标志变异指标有：全距、四分位间距、平均差及标准差、标志变动系数等。

1. 全距

全距（range）是指总体各单位标志值中最大数值与最小数值之差，又称极差，用“R”表示。它表明各个数值之间最大可能的差距，是一种测量标志变动度的最简单的方法。

对于未分组资料或单项数列资料， $R = \text{最大标志值} - \text{最小标志值}$

对于组距数列资料， $R \approx \text{最高组的上限} - \text{最低组的下限}$

【实例 5.17】 以实例 5.16 中男、女两参赛代表队成绩资料为例，求全距。

解：

男代表队全距： $R = 97 - 51 = 46$ （分）

女代表队全距： $R = 86 - 74 = 12$ （分）

男女两参赛代表队平均成绩都是 81 分，但从全距来看，男代表队 $R=46$ 分，表明成绩的差异大，所以平均数的代表性差；女代表队 $R=12$ 分，表明成绩的差异程度较小，所以用 81 分代表女队成绩，平均数的代表性较好。

全距主要用来说明总体各单位标志值变动总范围。全距越大，各标志值变动范围越大，平均数代表性较差；反之，全距越小，各标志值变动范围越小，平均数代表性就好；全距为 0 时，平均数的代表性最好。

全距的优点是计算方法简单，意义明确。全距经常用于检查产品质量的稳定性或进行产品质量控制。世界一些国家在提供证券市场行情时，广泛应用最高价、最低价和全距。

全距的缺点是易受极端数值的影响，它只能反映最大值和最小值之间的差距，而不能反映其内部各项数值的差异状况。当标志值有异常值存在时，会直接影响全距大小，使得全距不能充分反映总体分布的离散趋势。全距是一个较粗糙的测定标志变动度的指标。

2. 四分位数间距

四分位数间距（interquartile range）是将总体各单位标志值按一定顺序排列后，第三个四分位数与第一个四分位数之差，用“ IQR ”表示，即 $IQR=Q_3-Q_1$ 。第三个四分位数是处于中位数和最大数值中间的数值，第二个四分位数即中位数，第一个四分位数是处于中位数和最小数值中间的数值。

【实例 5.18】 根据美国普查局的资料，2014 年美国对前 15 个主要出口贸易伙伴的出口额资料如表 5.12 所示，求四分位数间距。

表 5.12 美国对前 15 个主要出口贸易伙伴的出口额资料表

项数序号	国家或地区	出口额（百万美元）
15	加拿大	312 125
14	墨西哥	240 326
13	中国	124 024
12	日本	66 964
11	英国	53 865
10	德国	49 443
9	韩国	44 544
8	荷兰	43 669
7	巴西	42 418
6	中国香港	40 877
5	比利时	34 824
4	法国	31 197
3	新加坡	30 532
2	中国台湾	26 836
1	澳大利亚	26 668

（注：项数按出口额由小至大排序。资料来源：全球经济数据 2015-02-11）

解：第一个四分位数的位置是 $\frac{1}{4} \times (15+1)=4$ ，所以： $Q_1=31\ 197$

第三个四分位数的位置是 $\frac{3}{4} \times (15+1)=12$ ，所以： $Q_3=66\ 964$

则四分位数间距 $IQR=Q_3-Q_1=66\ 964-31\ 197=35\ 767$ 。这表示，前 15 位中一半出口贸易伙伴出口额相差跨度不超过 35 767 百万美元。

四分位数间距先将数列一分为四，并略去最大与最小那一组的数值，然后再计算余下数值的全距。四分位数间距所包含的数值项数刚好是全部项数的一半。

四分位数间距的优点是克服了全距受极端数值影响的缺点，计算简便。在实际工作中，当人们对极端数值不感兴趣，而对处于当中部分的单位特征更在意时，常使用四分位数间距来计算。例如，在房地产行业中，用四分位数间距描述购房者中间的一半人，以及他们所购房屋价格的跨度。但是四分位数间距仍然不能反映总体内部各项标志值的差异状况，所以，它也是一个较粗糙的测定标志变动度的指标。

3. 平均差

全距与四分位数间距的大小都取决于数据中的两个数值。只有能够测度全部标志值对中心位置的偏差，才能对总体的离散程度做出最综合的说明。而要测度全部标志值偏离集中趋势的程度，很自然的设想就是计算各标志值偏离平均数的平均距离。由于各标志值对算术平均数的离差和为零，因此，最简单的做法就是将离差加上绝对值，然后求算术平均数。

平均差 (mean absolute deviation) 是总体各单位标志值与其算术平均数离差绝对值的算术平均数，用符号 “MAD” 表示。平均差表示总体分布中，各标志值对算术平均数的平均距离。平均差越大，说明各标志值差异程度越大，平均数代表性越差；平均差越小，说明各标志值差异程度越小，平均数代表性越好。

根据所掌握资料的不同，平均差可分为简单平均差和加权平均差两种。其基本公式为：

$$\text{简单平均差: } MAD = \frac{\sum |x - \bar{x}|}{n} \quad (\text{根据未分组资料计算时}) \quad (5-13)$$

$$\text{加权平均差: } MAD = \frac{\sum |x - \bar{x}| f}{\sum f} \quad (\text{根据分组资料计算时}) \quad (5-14)$$

【实例 5.19】 一家生产公司新引进一条产品生产线生产计算机，投产后头 5 个星期产量资料分别为 5 台、9 台、16 台、17 台、18 台，求平均差。

解：首先计算平均数：

$$\bar{x} = \frac{\sum x}{n} = \frac{5+9+16+17+18}{5} = 13$$

然后计算平均差：

$$MAD = \frac{\sum |x - \bar{x}|}{n} = \frac{|5-13| + |9-13| + |16-13| + |17-13| + |18-13|}{5} = 4.8 \text{ (台)}$$

【实例 5.20】 对某地区 90 家电子商务公司从业人员进行调查，员工人数资料如表 5.13 所示，求平均差。

表 5.13 90 家电子商务公司员工人数资料及计算表

员工人数	公司数 f	组中值 x	xf	$ x - \bar{x} $	$ x - \bar{x} f$
20 以下	32	10	320	32.89	1 052.48
20~40	16	30	480	12.89	206.24
40~60	13	50	650	7.11	92.43
60~80	10	70	700	27.11	271.1
80~100	19	90	1 710	47.11	895.09
合 计	90	—	3 860	—	2 517.34

解：首先计算平均数：

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{3\,860}{90} = 42.89$$

然后计算平均差：

$$MAD = \frac{\sum |x - \bar{x}|f}{\sum f} = \frac{2\,517.34}{90} = 27.97 \text{ (人)}$$

平均差是根据全部标志值计算出来的，反映了每个标志值与平均数的平均离散程度。与全距相比，受极端数值的影响较小，是比全距更优良的标志变异指标。在统计实践中，一般在产品质量控制分析中应用平均差。

但计算平均差时，须对离差取绝对值，这给平均差的代数运算带来了许多不便，从而使它的应用受到了限制。平均差并不是测定总体离散趋势的最好方法，在实际中，最常用的标志变异指标是标准差。

4. 标准差

标准差（standard deviation）是总体各单位标志值与其算术平均数离差平方的算术平均数的平方根，又称为均方差，用“ σ ”表示。标准差的平方称为方差（variance），用“ σ^2 ”表示。标准差是标志变异指标中最重要，也是最常用的指标。

标准差的实质与平均差基本相同，也表示各个标志值与平均数的平均离散程度。由于标准差采用平方而不是加绝对值的方法来消除各标志值与算术平均数离差的正负号，因此它的应用比平均差更广泛。

根据所掌握资料的不同，标准差可分为简单标准差和加权标准差两种。其基本公式为：

$$\text{简单标准差: } \sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} \quad (\text{根据未分组资料计算时}) \quad (5-15)$$

$$\text{加权标准差: } \sigma = \sqrt{\frac{\sum (x - \bar{x})^2 f}{\sum f}} \quad (\text{根据分组资料计算时}) \quad (5-16)$$

【实例 5.21】 根据实例 5.19 资料计算标准差。

解：（1）计算算术平均数：

$$\bar{x} = \frac{\sum x}{n} = \frac{5+9+16+17+18}{5} = 13$$

（2）计算各单位标志值与其算术平均数的离差：

“ $x - \bar{x}$ ”分别为-8、-4、3、4和5。

（3）计算各单位标志值与算术平均数的离差平方和：

$$\sum (x - \bar{x})^2 = 64 + 16 + 9 + 16 + 25 = 130$$

（4）计算离差平方和的算术平均数：

$$\frac{\sum (x - \bar{x})^2}{n} = \frac{130}{5} = 26$$

（5）开平方，得标准差：

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} = \sqrt{26} = 5.1 \text{ (台)}$$

实例 5.21 是资料未分组情况下简单标准差的计算。当资料分组时，算术平均数与标准差的计算均须采用加权平均法。加权平均法计算标准差的步骤与简单平均法基本相同。

【实例 5.22】 某地区 90 家电子商务公司员工人数资料如表 5.14 所示，计算电子商务公司员工人数标准差。

表 5.14 90 家电子商务公司员工人数标准差计算表

员工人数	公司数 f	组中值 x	xf	$x - \bar{x}$	$(x - \bar{x})^2$	$(x - \bar{x})^2 f$
20 以下	32	10	320	-32.89	1 081.752 1	34 616.067 2
20~40	16	30	480	-12.89	166.152 1	2 658.433 6
40~60	13	50	650	7.11	50.552 1	657.177 3
60~80	10	70	700	27.11	734.952 1	7 349.521
80~100	19	90	1 710	47.11	2 219.352 1	42 167.689 9
合 计	90	—	3 860	—	—	87 448.889

解：

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{3860}{90} = 42.89$$

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2 f}{\sum f}} = \sqrt{\frac{87448.889}{90}} = \sqrt{971.6543222} = 31.17 \text{ (人)}$$

因此, 这 90 家电子商务公司员工人数标准差为 31.17 人。

标准差是根据全部标志值计算的, 任何一个标志值发生变化, 都会使标准差发生变化。所以标准差反应灵敏, 能准确反映总体分布的离散趋势。另外, 它克服了平均差计算中用绝对值消除离差正负号带来的运算问题, 在计算上比平均差更简便。标准差是实际应用最广泛的标志变异指标。在统计分析中, 计算算术平均数的同时, 往往同时计算标准差。在相关、回归、抽样推断、统计预测等方面均要应用标准差。

但是由于标准差的大小与平均数有关, 且标准差带有与标志值相同的计量单位, 因此标准差不适宜用于比较平均数相差较大或计量单位不同的两总体分布的离散程度。

5. 标志变异系数

标志变异系数 (coefficient variation) 又称离散系数, 是用标志变异指标与相应的算术平均数对比, 来反映总体各单位标志值之间离散程度的指标, 是以相对数表示的标志变异指标。

前面介绍的各种标志变异指标都是表示标志变动程度的绝对指标, 其数值的大小, 还要受总体单位标志值本身水平高低的影响。在分析比较两组数字资料的离散趋势时, 如果两组数字的平均数相差较大, 或数字的计量单位不同, 则不能使用绝对变异指标, 必须进一步计算标志变异系数。标志变异系数有全距系数、平均差系数、标准差系数等, 实际统计分析工作中最常用的是标准差系数。

标准差系数的计算公式为:

$$CV = \frac{\sigma}{\bar{x}} \times 100\% \quad (5-17)$$

式中: CV——标准差系数。

【实例 5.23】 某学院会计专业 2015 级 2 个班, 英语期末考试平均成绩分别为: $\bar{x}_1=86$ 分, $\bar{x}_2=75$ 分; 标准差分别为: $\sigma_1=12$ 分, $\sigma_2=11$ 分。计算标准差系数, 并分析哪个班的成绩比较均匀, 哪个班平均成绩更具有代表性。

解:

$$\text{一班成绩的标准差系数 } CV_1 = \frac{\sigma}{\bar{x}} \times 100\% = \frac{12}{86} \times 100\% = 13.95\%$$

$$\text{二班成绩的标准差系数 } CV_2 = \frac{\sigma}{\bar{x}} \times 100\% = \frac{11}{75} \times 100\% = 14.67\%$$

虽然一班成绩的标准差大于二班成绩的标准差, 但由于两班平均成绩相差较大, 不能由此

说明一班平均成绩代表性比二班的差。为了说明两班考试成绩的离散程度，可以通过比较标准差系数来进行。由于一班成绩的标准差系数 13.95% 小于二班成绩的标准差系数 14.67%，所以得出结论：一班的考试成绩比较均匀，一班平均成绩更有代表性。

标准差系数是无名数，在应用时不受计量单位和标志值水平的限制，消除了不同总体之间在计量单位、平均水平方面的不可比性，适用于对比分析平均水平不同或计量单位不同的两组数据的离散程度的大小。标准差系数越大，说明数据的离散程度越大；标准差系数越小，说明数据的离散程度越小。

总之，反映总体离散趋势的各种标志变异指标，适用于不同的总体状况。其中，标准差应用最广。当需要对不同总体数据的离散程度进行比较时，标准差系数最常用。

5.5 Excel在平均指标和标志变异指标中的应用

5.5.1 Excel在平均指标中的应用

在 Excel 中，对未分组资料计算平均指标可利用 Excel 提供的相应函数计算；分组资料，则要借助于统计公式计算获得。

1. 数值平均数

数值平均数包括算术平均数、调和平均数和几何平均数，它们的操作步骤与情况类似。以算术平均数为例介绍数值平均数的计算。

计算算术平均数时，对于未分组资料，可用 AVERAGE() 计算平均数；对于分组资料，可用公式计算。

【实例 5.24】 利用第 3 章中【实例 3.4】提供的资料，计算 50 户居民购买消费品支出的平均数。

(1) 未分组资料。以【实例 3.4】的原始数据资料计算。未分组资料的计算有两种方法：其一，单击任一空单元格，输入 “=AVERAGE(A1:A50)”，回车得结果 1 196.2，这种方法比较简便直接；其二，在“插入”菜单中选择“函数”选项，在弹出的“粘贴函数”对话框的左侧“函数分类”中选择“统计”，在右侧的“函数名”中选择 AVERAGE，进入 AVERAGE 对话框，输入原始数据所在单元格区域，确定即可。AVERAGE 对话框如图 5.1 所示。

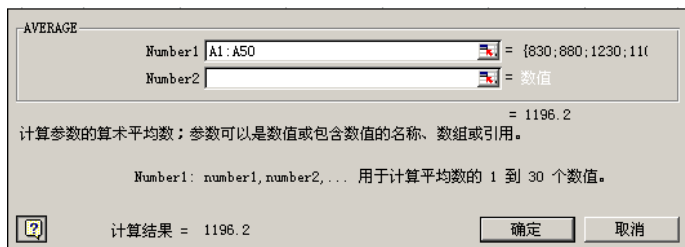


图 5.1 AVERAGE 函数对话框

(2) 分组资料。以【实例 3.4】的图 3.14 频数分布表的分组资料重新绘制表如下, 计算 50 户居民购买消费品支出的平均数, 如图 5.2 所示。

	A	B	C	D	E	F	G	H
	分组	频数f	组中值X	Xf				
1								
2	800~900	5	850	4250				
3	900~1000	1	950	950				
4	1000~1100	8	1050	8400				
5	1100~1200	11	1150	12650		1200		
6	1200~1300	11	1250	13750				
7	1300~1400	7	1350	9450				
8	1400~1500	4	1450	5800				
9	1500~1600	2	1550	3100				
10	1600以上	1	1650	1650				
11	合计	50		60000				
12								
13								

图 5.2 分组资料计算结果

具体操作步骤如下:

① 求出每组的组中值。

② 计算各组的消费品支出及支出总额, 单击 D2 单元格, 输入 “=B3*C2”, 回车得 D2 结果; 然后, 使用填充柄功能按住鼠标左键向下拖, 至 D10 单元格, 放开鼠标, 可得 D3~D10 结果。

③ 总额之和可利用 SUM 求和函数, 单击 D11 单元格, 输入 “=SUM(D2:D10)”, 回车得 D11 结果。

④ 在表外任一单元格, 输入 “=D11/B11”, 即可得平均数为 1 200。

在实际生活中, 直接利用调和平均数的情况非常少, 简单的调和平均数利用 Excel 提供的 HARMEAN()函数即可, 操作方法类似于未分组资料算术平均数的计算。间接地计算调和平均数利用公式计算, 操作方法类似于分组资料算术平均数的计算。

几何平均数适用于计算平均比率或平均速度的情况, 用 GEOMEAN()函数计算。

2. 位置平均数

位置平均数包括众数和中位数。

【实例 5.25】 利用第 3 章【实例 3.4】中提供的资料, 计算位置众数。

解: 使用未分组资料计算众数时用函数 MODE(), 操作方法类似于未分组资料算术平均数的计算; 计算分组资料的众数则利用众数的统计公式来计算。本例中众数所在组下限为 1 100 元, 众数组与前一组频数之差为 11-8=3, 众数组与后一组频数之差为 11-11=0, 众数组组距为 1 200-1 100=100。单击任一空单元格, 根据众数的下限公式, 输入 “=1100+(3/(3+0))*100”, 回车得结果 1 200。计算结果如图 5.3 所示。

The screenshot shows an Excel spreadsheet with a frequency distribution table in columns A and B. Column A is labeled '分组' (Group) and column B is labeled '频数f' (Frequency f). The data is as follows:

分组	频数f
800~900	5
900~1000	1
1000~1100	8
1100~1200	11
1200~1300	11
1300~1400	7
1400~1500	4
1500~1600	2
1600以上	1
合计	50

Formula bar: $=1100+(3/(3+0))*100$

图 5.3 众数计算结果

中位数的操作方法类似于众数。Excel 提供的中位数函数为 MEDIAN()。

5.5.2 Excel在标志变异指标中的应用

在 Excel 中，计算标志变异指标与计算平均指标类似，对未分组资料计算标志变异指标可利用 Excel 提供的相应函数计算；分组资料，则要借助于统计公式计算获得。

1. 全距

在 Excel 中求全距，分为下列两种情况。

(1) 资料未排序。如求【实例 3.4】的原始数据资料的极差，先利用 MAX()和 MIN()函数求得最大值和最小值。单击任一空单元格，此处为 B2，输入“=MAX(A1: A50)”，回车得最大值；再单击任一空单元格，此处为 B3，输入“=MIN(A1: A50)”，回车得最小值；最后单击任一空单元格，输入“=B2-B3”得全距为 820。

(2) 资料排序。用最大值和最小值直接相减可得全距。

2. 平均差

在 Excel 中求平均差，未分组资料，使用 AVEDEV()函数；分组资料，利用公式计算获得。

【实例 5.26】 利用【实例 3.4】提供的资料，计算平均差。

(1) 未分组资料。以【实例 3.4】的原始数据资料，计算其平均差。有两种方法：其一，单击任一空单元格，输入“=AVEDEV(A1:A50)”，回车得结果 147；其二，在“插入”菜单中选择“函数”选项，在弹出的“粘贴函数”对话框的左侧“函数分类”中选择“统计”，在右侧的“函数名”中选择 AVEDEV，进入 AVEDEV 对话框，输入原始数据所在单元格区域，确定即可得结果 147。

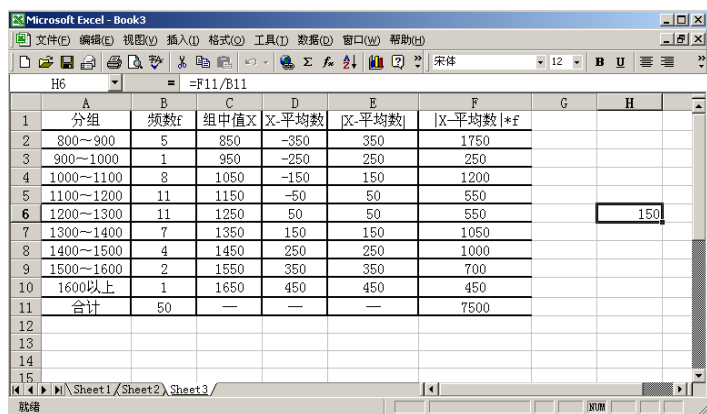
(2) 分组资料。以【实例 3.4】的频数分布表的分组资料（见图 3.14）计算其平均差，如图 5.4 所示。操作步骤如下。

第一步：计算出平均数，其计算参见本节平均数计算步骤。本例使用前面求得平均数 1 200 元。

第二步：计算各组购买消费品支出与平均购买消费品支出的离差的绝对值。本例，首先求得组中值与平均数的差，单击 D2 单元格，输入“=C2-1 200”，回车得 D2 结果；再使用填充柄功能按住鼠标左键向下拖，至 D10 单元格，放开鼠标，可得 D3~D10 结果。然后求离差的绝对值，单击 E2 单元格，输入“=ABS(D2)”，回车得 E2 结果；再使用填充柄功能按住鼠标左键向下拖，至 E10 单元格，放开鼠标，可得 E3~E10 结果。

第三步：计算离差的绝对值与频数的乘积。单击 F2 单元格，输入“=E2*B2”，回车得 F2 结果；再使用填充柄功能按住鼠标左键向下拖，至 F10 单元格，放开鼠标，可得 F3~F10 结果。乘积之和可利用 SUM 求和函数，单击 F11 单元格，输入“=SUM(F2:F10)”，回车得 F11 结果 7 500。

第四步：计算平均差。在表外任一单元格输入“=F11/B11”，即可得平均差为 150。



	A	B	C	D	E	F	G	H
	分组	频数f	组中值X	X-平均数	X-平均数	X-平均数 *f		
1								
2	800~900	5	850	-350	350	1750		
3	900~1000	1	950	-250	250	250		
4	1000~1100	8	1050	-150	150	1200		
5	1100~1200	11	1150	-50	50	550		
6	1200~1300	11	1250	50	50	550		150
7	1300~1400	7	1350	150	150	1050		
8	1400~1500	4	1450	250	250	1000		
9	1500~1600	2	1550	350	350	700		
10	1600以上	1	1650	450	450	450		
11	合计	50	—	—	—	7500		
12								
13								
14								
15								

图 5.4 分组资料计算平均差结果

3. 标准差与方差

标准差和方差是描述数据离散程度最常用的特征值。方差是标准差的平方。在 Excel 中求标准差与方差，未分组资料，使用函数；分组资料，利用公式计算获得。

【实例 5.27】 利用【实例 3.4】中提供的资料，计算标准差与方差。

(1) 未分组资料。在 Excel 中，用于计算标准差的函数有两个：一个是计算样本标准差的函数 STDEV()，另一个是计算总体标准差的函数 STDEVP()；相对应地，计算方差的函数也有两个：一个是计算样本方差的函数 VAR()，另一个是计算总体方差的函数 VARP()。

无论是样本资料还是总体资料，利用函数计算标准差和方差的方法同样有两种。

① 计算样本标准差和方差。【实例 3.4】中的原始资料是样本资料，计算其标准差与方差。

两种操作方法：其一，简便方法，单击任一空单元格，输入“=STDEV(A1:A50)”，回车得结果 188.65；其二，在“插入”菜单中选择“函数”选项，在弹出的“粘贴函数”对话框的左侧“函数分类”中选择“统计”，在右侧的“函数名”中选择 STDEV，进入 STDEV 对话框，

输入原始数据所在单元格区域，确定同样得结果 188.65。方差计算操作步骤类似，将函数换为样本方差函数 VAR 即可。

② 计算总体标准差与方差。假设【实例 3.4】中的原始资料是总体资料，计算其标准差与方差。

同样是两种操作方法：其一，简便方法，单击任一空单元格，输入“=STDEVP(A1:A50)”，回车得结果 186.75；其二，在“插入”菜单中选择“函数”选项，在弹出的“粘贴函数”对话框的左侧“函数分类”中选择“统计”，在右侧的“函数名”中选择 STDEVP，进入 STDEVP 对话框，输入原始数据所在单元格区域，确定同样得结果 186.75。

(2) 分组资料。以【实例 3.4】的频数分布表的分组资料（见图 3.14）计算其标准差和方差，如图 5.5 所示。具体操作步骤如下。

第一步：计算出平均数，其计算参见前述平均数的计算。本例使用前面求得平均数 1 200。

第二步：计算各组购买消费品支出与平均购买消费品支出的离差的平方。本例，首先求得组中值与平均数的差，单击 D2 单元格，输入“=C2-1 200”，回车得 D2 结果；再使用填充柄功能按住鼠标左键向下拖，至 D10 单元格，放开鼠标，可得 D3~D10 结果。然后求离差的平方，单击 E2 单元格，输入“=D2*D2”，回车得 E2 结果；再使用填充柄功能按住鼠标左键向下拖，至 E10 单元格，放开鼠标，可得 E3~E10 结果。

第三步：计算离差的平方与频数的乘积。单击 F2 单元格，输入“=E2*B2”，回车得 F2 结果；再使用填充柄功能按住鼠标左键向下拖，至 F10 单元格，放开鼠标，可得 F3~F10 结果。乘积之和可利用 SUM 求和函数，单击 F11 单元格，输入“=SUM(F2:F10)”，回车得 F11 结果 1 765 000。

第四步：计算方差和标准差。在表外任一单元格输入“=F11/B11”，本例中为 H5，可得方差为 35 300；再在表外任一单元格输入“=SQRT(H5)”，本例中为 H6，回车得标准差 187.88。或在表外任一单元格输入“=SQRT(F11/B11)”，可直接得标准差为 187.88。

	A 电子产品	B	C	D	E	F	G	H
	分组	频数f	组中值X	X-平均数	(X-平均数) ²	(X-平均数) ² *f		
1	800~900	5	850	-350	122500	612500		
2	900~1000	1	950	-250	62500	62500		
3	1000~1100	8	1050	-150	22500	180000		
4	1100~1200	11	1150	-50	2500	27500		35300
5	1200~1300	11	1250	50	2500	27500		187.8829
6	1300~1400	7	1350	150	22500	157500		
7	1400~1500	4	1450	250	62500	250000		
8	1500~1600	2	1550	350	122500	245000		
9	1600以上	1	1650	450	202500	202500		
10	合计	50	—	—	—	1765000		

图 5.5 分组资料计算标准差和方差的结果

4. 离散系数

在描述离散程度的各种特征值中，极差、平均差、方差和标准差都是绝对指标，各特征值的大小不仅受到离散程度的影响，而且还受到标志值本身的影响。因此，在对两个平均数不同的总体进行离散程度的比较时，使用标准差或方差是不适宜的，必须使用反映离散程度的相对数来比较，即使用离散系数来比较。

【实例 5.28】 有两组数据资料如下，分别计算其离散系数。

甲组：50 55 50 60 65 55 70 70 60 60

乙组：10 10 15 16 12 13 15 14 12 12

其操作步骤为：首先计算平均数，然后计算标准差，最后利用公式计算离散系数，本例计算结果如图 5.6 所示。简而言之，就是将前面的平均数与标准差的操作步骤相结合。

	A	B	C	D	E	F	G	H
1		甲组	乙组					
2		50	10					
3		55	10					
4		50	15					
5		60	16					
6		65	12					
7		55	13					
8		70	15					
9		70	14					
10		60	12					
11		60	12					
12	平均数	59.5	12.9					
13	标准差	6.873864	1.972308					
14	离散系数	0.115527	0.152892					
15								

图 5.6 离散系数计算结果

5.5.3 Excel描述统计工具应用

前面是 Excel 利用函数与公式来计算相应的特征值以描述数据的集中趋势与离散程度。对于统计数据的一些常用的统计量，如平均数、标准差等，Excel 提供了一种更加简便的方法——描述统计工具。用 Excel 描述统计工具可以同时给出平均数、标准误差、中位数（中值）、众数（模式）、样本标准差（标准偏差）、方差、峰度、极差（区域）、最小值、最大值、总和、观测数和置信度等十几个常用的统计量描述数据分布规律。以【实例 5.29】为操作范例介绍 Excel 描述统计工具的操作方法。

【实例 5.29】 从某高校经济管理系 2015 级 400 名学生中随机抽取 40 名学生，统计其英语成绩，得资料如下，用描述统计工具对大学生英语成绩进行分析。

69	61	75	60	80	72	82	86	90	78
62	87	71	77	79	87	85	97	85	63
94	78	65	85	96	88	89	82	87	78
95	83	56	81	58	75	78	74	67	75

解：

首先，在使用描述统计工具之前，先将样本数据排成一列并排序，本例中为 A1:A41。

然后，使用描述统计工具来操作，具体操作步骤如下。

第一步：单击“工具”菜单，选择“数据分析”选项。打开“数据分析”对话框，从其对话框的“分析工具”列表中选择“描述统计”，单击“确定”按钮后，打开“描述统计”对话框。

第二步：确定输入区域和输出选项，如图 5.7 所示。

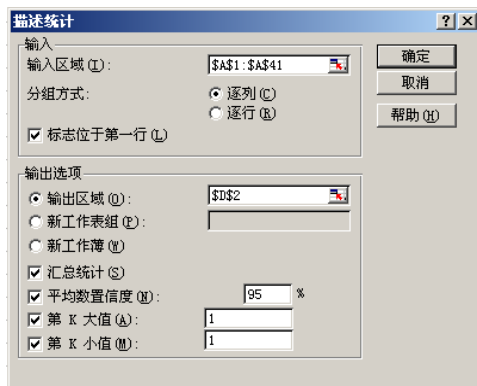


图 5.7 “描述统计”对话框

(1) 确定输入区域。

在“描述统计”对话框的“输入区域”框中输入分析数据所在的单元格区域。在本例中，输入区域为 \$A\$1:\$A\$41。分组方式中指出输入区域中的数据是按行还是按列排列，本例为“逐列”。若输入区域包括列标志行，则选中“标志位于第一行”复选框；如果输入区域无标志项，Excel 自动在输出区域加上“列 1”、“列 2”等作为标志。本例选中此复选框。

(2) 确定输出选项。

在“描述统计”对话框中可以指定结果的输出去向，共有三种。在“输出区域”框中输入输出结果所在的单元格区域，在本例中，输出区域为 \$D\$2。也可以通过选择“新工作表组”或“新工作簿”将结果放在新工作表或新工作簿中。

若选中“汇总统计”，则显示描述统计结果，否则不显示。一般为选中。

若选中“平均数值置信度”，则输出所包含均值的置信度。选中时需要再输入置信度，本例中输入的置信度为 95%，表明在显著性水平为 5%的条件下均值的置信度。

若选中“第 K 大值”或“第 K 小值”，则输出全数列中的 K 个最大值或最小值。选中时，需在右侧的文本框中输入 K 的数值，默认值为 1，即输出全数列的第 1 个最大值或最小值；若输入“2”，则输出全数列中的第 2 个最大值或最小值。本例中 K 值为 1，即输出第 1 个最大值和第 1 个最小值。

(3) 给出统计结果。单击“确定”按钮后，在指定位置给出描述统计结果，如图 5.8 所示。

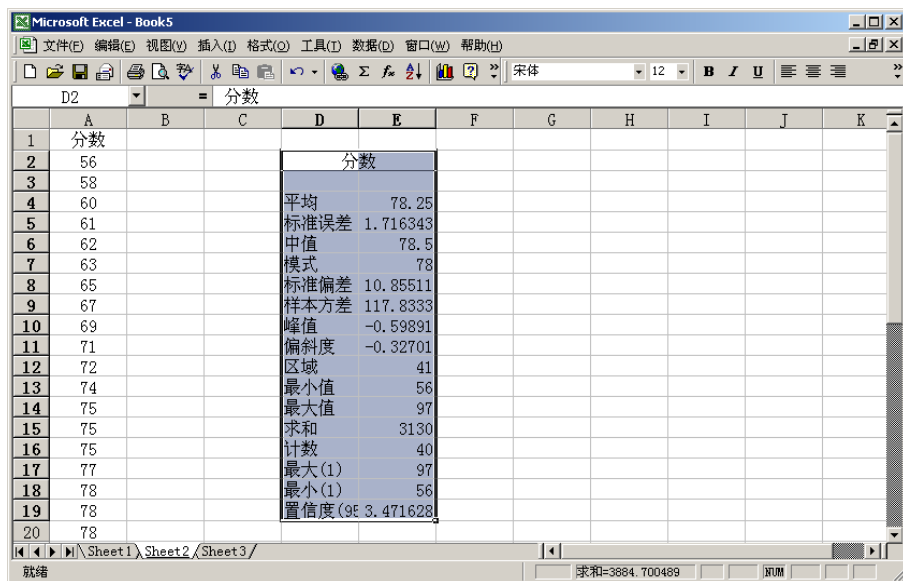


图 5.8 描述统计结果



统计术语

算术平均数 arithmetic mean

几何平均数 geometric mean

双众数 bimodal

中位数 median

全距 range

平均差 mean absolute deviation

标志变异系数 coefficient variation

变异程度测量 measures of variability

方差 variance

调和平均数 harmonic mean

众数 mode

多众数 multimodal

四分位数 quartiles

四分位数间距 interquartile range

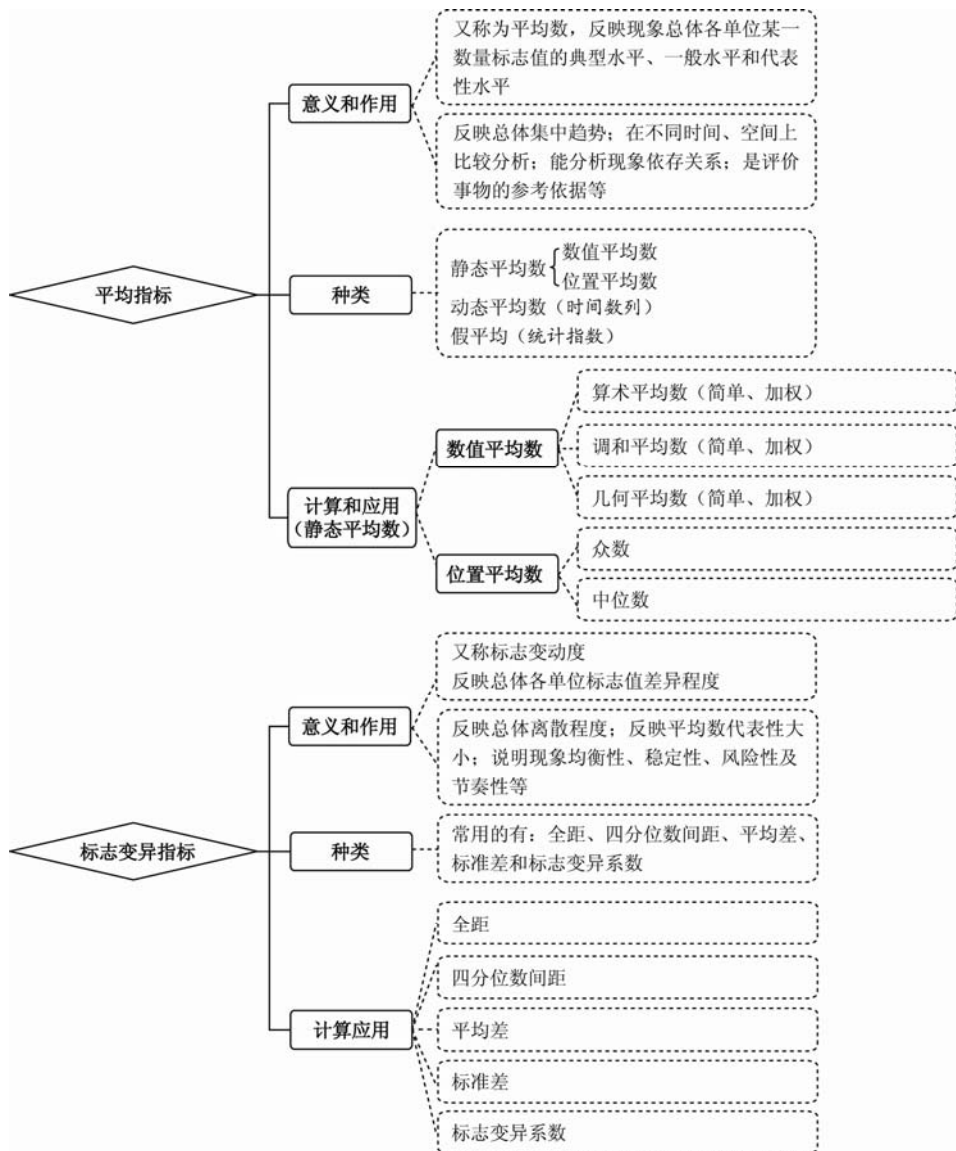
标准差 standard deviation

集中趋势测量 measures of central tendency

离差 deviation from the mean



重点知识梳理



习题与实践训练

一、判断题

- 2014 年按实际汇率计美国人均 GDP 和中国人均 GDP 分别是 57 045.46 美元和 8 211.49 美元。人均 GDP 是平均指标。 ()
- 对同一总体各单位标志值进行平均时，以次数为权数与以次数比重为权数，结果是不等的。 ()

- 3. 众数是总体中出现最多的次数。 ()
- 4. 简单算术平均数是加权算术平均数的特例，是权数相等时的加权算术平均数。 ()
- 5. 中位数和众数都属于平均数，因此它们数值的大小受到总体各单位标志值大小的影响。 ()
- 6. 标准差系数越大，说明总体中各单位标志值的变异程度越大，则平均指标的代表性就越好。 ()
- 7. 在对比两个不同总体各单位标志值的离散程度时，应采用标准差而不是全距或平均差指标进行比较。 ()
- 8. 中位数就是处于中间的第二个四分位数。 ()
- 9. 数值平均数属于静态平均数，而位置平均数属于动态平均数。 ()
- 10. 将全部标志值分成两半，一半小于中位数，一半大于中位数。 ()

二、单项选择题

- 1. 在分配数列中，当标志值较小而权数较大时，计算的算术平均数 ()。
 - A. 接近于标志值较大的一方
 - B. 接近于标志值较小的一方
 - C. 接近于大小合适的标志值
 - D. 不受权数影响
- 2. 若根据同一分组资料计算简单算术平均数和加权算术平均数，结果相同，则可推定 ()。
 - A. 各组权数相等
 - B. 各组权数不等
 - C. 变量值大致相等
 - D. 计算错误
- 3. 标志变异指标 ()。
 - A. 说明总体各单位标志值的变异程度
 - B. 把总体各单位标志值的差异抽象化了
 - C. 能够反映总体分布的集中趋势
 - D. 与平均数的代表性成正比
- 4. 在不掌握各组单位数资料，只掌握各组标志值和各组标志总量的情况下，宜采用 ()。
 - A. 加权算术平均数
 - B. 几何平均数
 - C. 加权调和平均数
 - D. 简单算术平均数
- 5. 标志变异指标中易受极端数值影响的是 ()。
 - A. 全距
 - B. 平均差
 - C. 标准差
 - D. 标准差系数
- 6. 某市 2015 年农村人均收入和城市人均收入分别为 9 600 元和 30 180 元，标准差分别为 640 元和 2 340 元，则人均收入的变异程度 ()。
 - A. 城市大
 - B. 一样大
 - C. 农村大
 - D. 不可比
- 7. 随机调查某城市 100 户家庭，得家庭订阅报纸杂志份数资料如下表所示。

报纸杂志数	0	1	2	3	4	5	合 计
家庭户数	9	54	21	12	2	2	100

- 则根据这份资料计算的众数是 ()。
- A. 21
 - B. 1
 - C. 54
 - D. 2
8. 在甲乙两个变量数列中，若甲数列标准差 σ_1 <乙数列标准差 σ_2 ，则两个变量数列平均数代表性相比较 ()。

- A. 甲数列的平均数代表性高于乙数列
 - B. 两个数列的平均数代表性相同
 - C. 乙数列的平均数代表性高于甲数列
 - D. 不能确定
9. 两个总体平均数不等, 但标准差相等, 则有 ()。
- A. 两个平均数代表性相同
 - B. 平均数大代表性小
 - C. 平均数小代表性小
 - D. 无法进行正确判断
10. 下列指标中属于位置平均数的有 ()。
- A. 算术平均数
 - B. 四分位数间距
 - C. 四分位数
 - D. 全距

三、多项选择题

1. 在平均指标中, 要根据所有标志值计算的平均数是 ()。
- A. 中位数
 - B. 算术平均数
 - C. 众数
 - D. 几何平均数
 - E. 调和平均数
2. 下列属于平均指标的有 ()。
- A. 人口密度
 - B. 某公司职工平均年龄
 - C. 参加英语等级考试的学生的平均成绩
 - D. 全国人均钢产量
 - E. 居民人均可支配收入
3. 与标志值有相同计量单位的标志变异指标是 ()。
- A. 全距
 - B. 平均差
 - C. 四分位数间距
 - D. 标准差
 - E. 标准差系数
4. 中位数是 ()。
- A. 由标志值在数列中的位置决定的
 - B. 根据标志值出现的次数决定的
 - C. 总体一般水平的代表值
 - D. 不受总体中极端数值的影响
 - E. 总体各单位平均指标
5. 当一个变量数列中出现个别极端数值时, 这些极端值 ()。
- A. 对算术平均数、众数、中位数均有影响
 - B. 对算术平均数、众数、中位数都没有影响
 - C. 对众数无影响, 对算术平均数有影响
 - D. 对众数和中位数均无影响
 - E. 对算术平均数影响最大, 对中位数影响最小
6. 两个总体平均数相等, 但标准差不等, 则 ()。
- A. 标准差越大, 平均数代表性越小
 - B. 标准差越小, 各单位标志值差异程度越小
 - C. 两总体集中趋势相同
 - D. 无法比较两总体的离散程度
 - E. 无须计算标准差系数, 直接用标准差比较两总体离散趋势

7. 在 () 的条件下, 加权算术平均数等于简单算术平均数。
 - A. 各组次数均为 1
 - B. 各组次数占总次数的比重相等
 - C. 各组次数相等
 - D. 各组变量值不等
 - E. 变量数列为组距数列
8. 两种价格不同的食品, 在销售量都增加的情况下, 两种食品的平均价格 ()。
 - A. 肯定变动
 - B. 肯定不变
 - C. 可能变动也可能不变
 - D. 若销售量成比例增加, 则平均价格不变
 - E. 若销售量增加数相同, 则平均价格不变
9. 下列可应用加权算术平均法计算平均数的有 ()。
 - A. 由各员工的工资额求平均工资
 - B. 由各员工按工资分组的变量数列求平均工资
 - C. 由工资总额及员工总数计算平均工资
 - D. 由各道工序的合格率求产品的平均合格率
 - E. 由每一年的经济发展速度求 5 年间平均发展速度
10. 平均指标与标志变异指标的关系是 ()。
 - A. 平均指标是对总体各单位标志值一般水平的测度, 代表程度取决于标志变异指标的大小
 - B. 标志变异指标越大, 平均指标代表性越小
 - C. 标志变异指标越小, 平均指标代表性越好
 - D. 平均指标和标志变异指标分别反映同一总体的集中趋势和离散趋势
 - E. 两者无关系

四、填空题

1. 平均指标反映了总体各单位某一数量标志值的_____。
2. 权数对算术平均数的影响作用, 就其实质来看, 不是决定于_____的多少, 而是决定于_____的大小。
3. 已知三种产品的合格率分别为 58%、76%、91%, 则三种产品的平均合格率为_____。
4. 已知某产品流水线上三道工序产品合格率分别为 94%、91%、97%, 则该产品平均合格率为_____。
5. 直接用平均差和标准差比较两个变量数列平均数的代表性的前提条件是两个变量数列的_____相等。
6. 众数是在总体中出现次数_____的标志值; 中位数是位于数列_____位置的标志值。中位数和众数也可称为_____平均数。
7. 算术平均数是_____与_____之比。
8. 众数不受资料中_____的影响, 如果变量数列中_____, 则没有众数。
9. 某班级统计学考试平均成绩为 85 分, 标准差为 10 分, 则标准差系数是_____。
10. 平均指标反映总体数据的_____趋势; 标志变异指标反映总体数据的_____程度。

五、应用能力训练题

1. 下表是美国《财富》杂志于 2015 年 7 月 22 日公布的 2015 年世界 500 强企业中前 15 强企业的营业收入和利润资料，请求出营业收入的平均数和中位数。

排名	公司名称	营业收入（百万美元）	利润（百万美元）	国家
1	沃尔玛	485 651	16 363	美国
2	中国石油化工集团公司	446 811	5 177	中国
3	荷兰皇家壳牌石油公司	431 344	14 874	荷兰
4	中国石油天然气集团公司	428 620	16 359.5	中国
5	埃克森美孚	382 597	32 520	美国
6	英国石油公司	358 678	3 780	英国
7	国家电网公司	339 426.5	9 796.2	中国
8	大众公司	268 566.6	14 571.9	德国
9	丰田汽车公司	247 702.9	19 766.9	日本
10	嘉能可	221 073	2 308	瑞士
11	道达尔公司	212 018	4 244	法国
12	雪佛龙	203 784	19 241	美国
13	三星电子	195 845.3	21 922.7	韩国
14	伯克希尔-哈撒韦公司	194 673	19 872	美国
15	苹果公司	182 795	39 510	美国

2. 某集团下属 20 个公司在 2016 年第一季度产值计划完成情况如下表所示。

计划完成程度（%）	公司个数	计划任务数
90~100	3	80
100~110	12	400
110~120	5	120
合 计	20	

试计算平均计划完成程度。

3. 某市一个农贸市场 2016 年 1 月三种蔬菜的价格、成交额资料如下表所示。

蔬菜名称	价格（元/千克）	成交额（元）
土豆	2.4	2 400
油菜	3.2	6 400
蘑菇	4.4	4 400
合 计	—	13 200

计算蔬菜的平均价格。

4. 某学院工商管理系共有 120 名学生选修统计学，在期末考试中，男生平均成绩为 77 分，女生平均成绩为 81 分。则：

（1）若 120 名学生中，男、女生各占一半，全体学生平均成绩为多少？

- (2) 若 120 名学生中, 男生 80 人, 女生 40 人, 全体学生平均成绩为多少?
- (3) 若 120 名学生中, 男生 40 人, 女生 80 人, 全体学生平均成绩为多少?
- (4) 比较上述三种情况下, 全体学生平均成绩的变化, 解释变化的原因。

5. 某市广播电台刚开办了《经典音乐回放》栏目, 想了解听众年龄特征, 以决定播放时段。随机抽查收听该节目的 170 个听众, 得年龄分布资料如下表所示。

年龄 (岁)	听众人数 (人)
15~20	9
20~25	16
25~30	27
30~35	44
35~40	42
40~45	23
45~50	7
50~55	2
合 计	170

- (1) 听众年龄的平均数为多少? 众数为多少?
- (2) 计算听众年龄的标准差。

6. 中国互联网络中心 (CNNIC) 2015 年 7 月发布第三十六次中国互联网络发展状况统计报告显示, 中国互联网用户年龄 2015 年 6 月分布资料如下。

年龄 (岁)	10 岁以下	10~19 岁	20~29 岁	30~39 岁	40~49 岁	50~59 岁	60 岁及以上
比重 (%)	1.8	23.8	31.4	23.2	13.0	4.3	2.4

请计算中国互联网用户平均年龄、标准差。

7. 一个总体各单位标志值的平均数是 12.3 分钟, 标准差是 4 分钟。

- (1) 如果给每个标志值都加上 5 分钟, 平均数是多少?
- (2) 如果给每个标志值都加上 5 分钟, 标准差是多少?
- (3) 如果给每个标志值都加上 5 分钟, 方差是多少?
- (4) 如果给每个标志值都加上 5 分钟, 众数、中位数、四分位数是否发生变化, 为什么?

8. 一种儿童玩具需要人工组装, 现有三种可供选择的组装方法。为检验哪种方法更好, 随机抽取 15 名工人, 让他们分别用三种方法组装。15 名工人分别用三种方法在相同的时间内组装的产品数量如下。

单位: 个															
第一种方法	127	129	129	130	130	130	129	131	127	128	128	128	128	132	125
第二种方法	162	164	165	167	168	164	165	168	170	163	164	166	165	166	167
第三种方法	162	168	169	169	170	152	150	169	174	170	161	169	160	171	167

请运用本章所学知识, 说明采用什么指标评价组装方法的好坏? 如果让你选择一种方法, 你将选择第几种方法? 请计算并说明理由。

9. 对 10 名成年人和 10 名幼儿的身高进行抽样调查，得到如下资料。

幼儿组身高（cm）	74	70	68	71	69	75	72	68	73	73
成年组身高（cm）	168	166	169	180	177	172	170	172	174	173

试比较：成年组和幼儿组哪一组的身高差异大。

10. 根据第 35 次中国互联网络发展状况统计报告，2013—2014 年中国内地分省网民规模及互联网普及率资料如下。

省份	网民数（万人）	普及率	网民规模增速	普及率排名
北京	1 593	75.3%	2.4%	1
上海	1 716	71.1%	2.0%	2
广东	7 286	68.5%	4.2%	3
福建	2 471	65.5%	2.9%	4
浙江	3 458	62.9%	3.9%	5
天津	904	61.4%	4.4%	6
辽宁	2 580	58.8%	5.2%	7
江苏	4 274	53.8%	4.4%	8
山西	1 838	50.6%	4.7%	9
新疆	1 139	50.3%	4.2%	10
青海	289	50.0%	5.5%	11
河北	3 603	49.1%	6.3%	12
山东	4 634	47.6%	7.0%	13
海南	421	47.0%	2.3%	14
陕西	1 745	46.4%	3.3%	15
内蒙古	1 142	45.7%	4.5%	16
重庆	1 357	45.7%	4.9%	17
湖北	2 625	45.3%	5.4%	18
吉林	1 243	45.2%	6.9%	19
宁夏	295	45.1%	4.2%	20
黑龙江	1 599	41.7%	5.6%	21
西藏	123	39.4%	6.9%	22
广西	1 848	39.2%	4.2%	23
湖南	2 579	38.6%	7.0%	24
四川	3 022	37.3%	6.6%	25
河南	3 474	36.9%	5.8%	26
安徽	2 225	36.9%	3.5%	27
甘肃	951	36.8%	6.4%	28
云南	1 643	35.1%	7.5%	29
贵州	1 222	34.9%	6.7%	30
江西	1 543	34.1%	5.1%	31
全国	64 875	47.9%	5.0%	--

通过变异系数来反映省间互联网普及率的差异，可以看到，我国互联网普及的地区差异呈现稳定的下降趋势，截至 2014 年 12 月，互联网普及率的省间差异为 0.24，相比 2013 年年底下降了 0.01。实现互联网接入以来，中国在推进互联网全面普及的工作上取得显著成效，但由于互联网普及率的省间差异仍然存在，进一步推动普及情况落后省份的互联网建设工作将成为一项长期工程。（摘自《第 35 次中国互联网络发展状况统计报告》腾讯科技 2015-02-03）

请计算我国互联网普及率的平均数、中位数及标准差系数。

11. 从生产车间 500 名工人中随机抽取 50 名工人，50 名工人的日产量（件）如下所示。

148 140 127 120 110 104 132 135 129 123
116 109 132 135 129 123 110 108 148 135

128	123	114	108	132	124	120	125	116	118
125	137	107	113	123	140	137	119	119	127
128	119	124	130	118	107	113	122	128	114

请进行以下统计分析：

- (1) 用 Excel 函数计算平均数、众数、中位数、极差、平均差、标准差。
- (2) 利用分组资料计算平均数、众数、中位数、极差、平均差、标准差。



本章案例

北京市统计局 2015 年 6 月 5 日公布，2014 年度全市职工平均工资为 77 560 元，比上年增长 11.6%；月平均工资为 6 463 元，比 2013 年的 5 793 元多了 670 元。全市职工平均工资与多项社会保障项目直接挂钩，2014 年度全市职工平均工资计算的事项均将按“新标准”执行。数据显示，2014 年城镇非私营单位比私营单位员工年均工资高了近 5 万元，差距比上年扩大 4 387 元。年平均工资最高的行业依然是金融业，为 225 482 元，是全市平均水平的 2.2 倍。

2014 年，市统计局首次发布分岗位平均工资数据，以反映不同岗位的工资差距。2014 年调查显示，调查单位就业人员平均工资为 90 900 元，不同岗位平均工资差距缩小。从岗位工资统计总体情况看，中层及以上管理人员的平均工资最高，商业、服务业人员平均工资最低，最高与最低工资之比为 4.36 : 1，与 2013 年的 4.40 : 1 相比有所缩小。

城镇非私营单位中，年平均工资最高的三个行业为金融业，信息传输、软件和信息技术服务业，卫生和社会工作。其中金融业年平均工资 225 482 元，是全市平均水平的 2.2 倍。年平均工资最低的行业是居民服务、修理和其他服务业，仅为全市平均水平的 44.8%。

市统计局有关负责人表示，平均工资数据反映的是全市城镇非私营、私营单位就业人员的平均水平。由于工资分布是典型的偏态分布，即少数人工资水平较高，多数人工资水平较低，所以多数人的工资水平会低于平均工资。同时，由于单位所处的行业、隶属关系、单位性质、经济效益及个人所在的岗位不同等诸多因素，工资水平客观上存在较大差异，每个人对平均工资增长的感受也不尽相同。

2014 年本市城镇非私营单位就业人员年平均工资为 102 268 元，2014 年本市城镇私营单位就业人员年平均工资为 52 902 元。

(资料来源：北京晨报 2015-06-06)

抽样推断

学习要点

- 抽样推断的基本概念。
- 简单随机抽样形式的区间估计方法。
- 必要样本单位数的确定方法。
- 假设检验的基本理论和步骤。
- 总体均值和总体成数的假设检验方法。

抽样法是从全部研究对象中抽取一部分对象进行观察，并以此来对全部研究对象做出判断。在我们日常生活中，随处可见抽样法的应用。例如，我们到集市上去买花生、瓜子总是要抓几粒看看是否饱满；在家里做菜，做好后要尝尝咸淡如何；生病到医院看病时，医生抽一些血来化验，等等。在统计的搜集资料和分析研究中也常常要用到抽样的方法，利用这种方法来搜集资料就称作抽样调查。

科学的抽样调查并不像上面所举的日常生活中的例子那样简单，因为上面所举的例子中可以认为研究的全部对象中各部分的分布是比较均匀的。例如，在炒瓜子、花生和做菜过程中都是搅拌均匀，人体的血液循环过程中是分布均匀，因此，抽取少量样品就有很强的代表性。但是在社会经济的调查中，研究对象的差异往往比较大，如居民家庭收支情况受到地区、职业、工种及家庭就业人员数等许多因素的影响，差别很大；农作物的产量则受土壤、肥料及管理等多种因素的影响，各个地块的产量也高低不一，那么要使抽选出来的家庭或地块对全体有代表性就不那么容易了。另外，抽样调查以一部分来推断全体，也难免出现代表性误差。如何计量和控制这些误差，使我们能以一定的准确度和把握度（可靠度）利用样本信息推断总体特征，这就是抽样推断所要研究解决的问题。

根据研究问题的侧重点和出发点不同，统计推断主要分为参数估计和参数假设检验两大部分。例如，某外贸公司从国外进口一批产品须通过抽取样本检验其产品是否符合合同要求。如果是根据样本算出的合格率，以一定的把握程度推算出全批产品的合格率数值范围，这就是参数估计问题；如果是根据合同要求，事先规定合格率低于某一具体数值就不接受，然后利用样本资料对这批产品质量标准进行推断，这就是假设检验问题。

6.1 抽样推断的基本概念

抽样推断 (sampling inference) 是在抽样调查 (sampling survey) 的基础上, 利用样本的实际资料计算样本指标, 并据此推算总体相应数量特征的一种统计分析方法。在第4章和第5章中, 我们已经学习了计算各项综合指标, 如总量指标、相对指标、平均指标等, 来反映总体的数量特征。但是, 在现实生活中面对经济总体时, 常存在如下三种情况: 一是不可能对总体全面调查。当所研究对象为无限总体, 或研究目的是寿命测试从而导致研究具有破坏性时, 我们不可能逐一进行调查。二是来不及对总体进行全面调查。流水线上产品的质量检查、所有商品的进出口检验等, 由于数量巨大, 时间要求受限, 来不及进行全面调查。三是没有必要对总体全面调查, 如民意测验、城乡居民的电视收视率的调查等。上述情形普遍存在于社会经济生活中, 我们只能组织抽样调查, 取得部分的实际资料, 来估计和推断总体的数量特征, 以达到对现象总体的认识。

归纳起来, 抽样推断具有四个特点: 第一, 抽样推断是由部分推算整体的一种认识方法。在第2章统计调查中我们已讲到, 抽样调查是一种非全面调查方式, 其调查目的并不是了解部分单位的情况, 而是作为认识总体的一种手段。利用样本的部分信息, 推断总体的相关信息。第二, 抽样推断是建立在随机取样的基础上。抽样调查可以是概率抽样, 也可以是非概率抽样, 但是作为抽样推断基础的必须是概率抽样, 即按随机原则抽取样本单位, 这是抽样推断的前提。这样才能保证样本变量是随机变量, 抽出的样本才具有代表性。第三, 抽样推断是运用概率估计的方法, 利用样本指标来估计总体参数。在数学上运用不确定的概率估计法, 而不是运用确定的自变量和因变量的关系, 它不能利用一定的函数关系来推算总体参数。原则上把由样本观察值所决定的样本指标看作随机变量。因此, 用样本指标值推断总体指标值, 是以一定可靠性作保证的。第四, 抽样推断的误差是可以预先控制的, 用样本指标值推断总体指标值是存在一定误差的。但是, 由于抽样推断是有科学根据的, 所以, 抽样误差的范围可以事先通过有关资料加以计算, 并且加以控制, 从而使抽样推断具有一定的可靠度, 这是其他估算方法所不及的。抽样推断有其专属的基本概念, 它们是理解和进行抽样推断研究的基础。

6.1.1 总体和样本

1. 总体

总体 (population) 也称全及总体, 是指所研究现象的整体, 即包括所要调查的所有单位。例如, 从1 000名学生中抽取100名学生进行抽样调查, 以计算学生的平均体重。这1 000名学生是全及总体, 一般用英文大写字母“ N ”来表示总体的单位数, 取 $N=1\ 000$ 人。全及总体按其各单位性质的不同, 可以分为变量总体和属性总体两类。对于一个总体来说, 若被研究的单位标志属于品质标志, 则该总体为属性总体; 若被研究的单位标志属于数量标志, 则该总体为变量总体。

2. 样本

样本 (sample) 又称子样, 它是从全及总体中随机抽取出来, 作为代表这一总体的那部分单位组成的集合体。样本的单位数是有限的, 相对来说, 它的数目比较小, 一般用英文小写字母 “ n ” 来表示样本的单位数, 如 $n=100$ 人。

作为推断对象的总体是确定的, 而且是唯一的, 但作为观察对象的样本就不是这样。从一个总体可以抽取很多个样本, 每次可能抽到哪个样本不是确定的, 也不是唯一的, 而是可变的、随机的。

6.1.2 参数和统计量

1. 参数 (parameter)

根据总体各单位的标志值或标志属性计算的, 并反映总体数量特征的综合指标称为全及指标。全及指标是总体变量的函数, 其数值是由总体各单位的标志值决定的。由于总体是唯一确定的, 因此, 全及指标也是唯一确定的, 是不变的, 所以也称参数。常用的参数有总体平均数、总体成数、总体方差和总体标准差。

(1) 总体平均数。对于变量总体, 由于各单位的标志用数量表示, 因此可以计算总体平均数, 通常用 \bar{X} 表示。

设 X 为总体的某一变量, 其 N 项变量值为 X_1, X_2, \dots, X_n , 则总体平均数为:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{N} = \frac{\sum X}{N} \quad (6-1)$$

(2) 总体成数。对于属性总体, 由于各单位标志不能用数量表示, 因此总体参数常以成数或比重来表示。通常以 P 表示总体中具有某种标志表现, 即 “是” 的单位数在总体单位数中所占的比重; 以 Q 表示不具有某种标志表现, 即 “非” 的单位数所占的比重。

设总体 N 个单位中, 有 N_1 个单位具有某种标志表现, N_0 个单位不具有某种标志表现, 且 $N=N_1+N_0$, 则总体成数为:

$$P = \frac{N_1}{N} \quad (6-2)$$

$$Q = \frac{N_0}{N} = \frac{N - N_1}{N} = 1 - P \quad (6-3)$$

【实例 6.1】 红光灯泡厂生产的 10 000 只灯泡中, 合格品有 9 550 只, 不合格产品有 450 只, 则总体成数即灯泡的合格率和不合格率分别为:

$$P = \frac{N_1}{N} = \frac{9\,550}{10\,000} = 95.5\%$$

$$Q = \frac{N_0}{N} = \frac{450}{10\,000} = 4.5\%$$

需要指出，统计上把只表现为“是”与“非”的标志称为是非标志，也称交替标志（如性别标志等）。

(3) 总体方差和总体标准差。就变量总体而言，总体方差和标准差计算公式分别为：

$$\sigma^2 = \frac{\sum (X - \bar{X})^2}{N} \tag{6-4}$$

$$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N}} \tag{6-5}$$

在属性总体条件下，则可以把“是”与“非”两种标志表现进行量化处理。用“1”表示“是”，即具有某种标志表现；用“0”表示“非”，即不具有某种标志表现。那么“是”的成数就可视为是非标志的平均数，从而计算出属性总体的方差和标准差，即：

$$\sigma_p^2 = (1 - P)^2 P + (0 - P)^2 Q = Q^2 P + P^2 Q = PQ(P + Q) = PQ = (1 - P) \tag{6-6}$$

$$\sigma = \sqrt{P(1 - P)} \tag{6-7}$$

属性总体的方差和标准差的具体计算过程如表 6.1 所示。

表 6.1 属性总体的方差和标准差的计算表

变 量	成数或频率	变量与频率之积	离差平方	离差平方与频率之积
X	$\frac{f}{\sum f}$	$X \square \frac{f}{\sum f}$	$(X - P)^2$	$(X - P)^2$
1 (是)	P	P	$(1 - P)^2$	$(1 - P)^2 P$
0 (非)	Q	0	$(0 - P)^2$	$(0 - P)^2 Q$
合 计	1	P	—	$P(1 - P)$

【实例 6.2】 仍以实例 6.1 的有关资料为例，计算其总体方差和标准差分别为：

$$\sigma_p^2 = P(1 - P) = 95.5\% \times (1 - 95.5\%) = 4.3\%$$

$$\sigma_p = \sqrt{P(1 - P)} = \sqrt{95.5\% \times (1 - 95.5\%)} = 20.73\%$$

在抽样调查中，总体参数的意义和计算方法是明确的，但参数的具体数值事先是未知的，需要用抽样来估计它。

2. 统计量 (statistic)

根据样本各单位标志值计算出的反映样本特征的指标称为统计量，也称作样本指标，它是用来估计总体参数的。与总体参数相对应，统计量主要有样本平均数、样本成数、样本方差和样本标准差等。

(1) 样本平均数。设 x 为样本的某一变量，其 n 项变量值为 x_1, x_2, \cdots, x_n ，则样本平均数为：

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum x}{n} \quad (6-8)$$

式中: \bar{x} —— 样本平均数。

(2) 样本成数。设样本的 n 个单位中有 n_1 个单位具有某种标志表现, n_0 个不具有某种标志表现, 且 $n=n_1+n_0$, 则样本成数为:

$$p = \frac{n_1}{n} \quad (6-9)$$

$$q = \frac{n_0}{n} = \frac{n - n_1}{n} = 1 - p \quad (6-10)$$

式中: p 、 q —— 样本成数。

【实例 6.3】 从红光灯泡厂生产的灯泡中, 抽样检查了 100 只灯泡, 其中, 合格品 95 只, 不合格品 5 只, 则样本成数即样本的合格率和不合格率分别为:

$$p = \frac{n_1}{n} = \frac{95}{100} = 95\%$$

$$q = \frac{n_0}{n} = \frac{5}{100} = 5\%$$

(3) 样本方差和样本标准差。它们是说明样本内各单位标志值变异程度的指标。

样本数量标志的方差 S^2 和标准差 S 的计算公式分别为:

$$S^2 = \frac{1}{n} \sum (x - \bar{x})^2 \quad (6-11)$$

$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} \quad (6-12)$$

样本是非标志的方差 S^2 和标准差 S 的计算公式分别为:

$$S_p^2 = P(1 - P) \quad (6-13)$$

$$S_p = \sqrt{P(1 - P)} \quad (6-14)$$

【实例 6.4】 以实例 6.3 有关抽样调查资料为例, 可计算出样本方差和标准差分别为:

$$S_p^2 = P(1 - P) = 95\% \times (1 - 95\%) = 4.75\%$$

$$S_p = \sqrt{P(1 - P)} = \sqrt{95\% \times (1 - 95\%)} = 21.79\%$$

统计量作为样本变量的函数, 其计算方法是确定的, 但它的取值会随着样本的不同而变化, 所以统计量本身也是随机变量。

6.1.3 样本容量和样本个数

样本容量和样本个数是两个互相联系但又完全不同的概念。样本容量是指一个样本所包含的单位数。一个样本应该包含多少单位最合适,是抽样设计必须认真考虑的问题。必须结合调查任务的要求及总体标志值的变异情况来考虑。样本容量的大小不但关系到抽样调查的效果,而且关系到抽样方法的应用。通常将样本单位数不少于 30 个的样本称为大样本,单位数不及 30 个的样本称为小样本。社会经济统计的抽样调查多属于大样本调查。

样本个数又称样本可能数目,是指从一个总体中可能抽取的样本个数。一个总体可能抽取多少样本,和样本容量及抽样方法等因素都有关系,是一个比较复杂的问题。一个总体有多少样本,样本统计量就有多少种取值,从而形成统计量的分布。统计量的分布是抽样推断的基础。虽然在实践上只抽取个别或少数样本,但要判断所取样本的可能性就必须联系到全部可能取样本数目所形成的分布。

6.1.4 重复抽样和不重复抽样

重复抽样又叫重置(with replacement)抽样,是指从所研究的现象总体中,按照随机原则,抽取一个样本单位进行调查登记之后,把这个样本再放回总体中去,然后再从所研究的总体中抽取样本单位。这样,总体单位数不变,设它为 N 。在进行第二次和第三次抽取样本时,已经被抽中的样本仍然有同等的机会再被抽中。

不重复抽样又叫不重置(without replacement)抽样,是指从所研究的现象总体中,按照随机原则,抽取一个样本单位进行调查登记之后,不再把这个样本放回去。这样,所研究现象的总体单位数将逐步减少,即总体单位数在不断起变化。

由于重复抽样和不重复抽样在抽样方法上不同,抽取的样本个数也不相同。从总体 N 个单位中,用重复抽样的方法,随机抽样取 n 个单位构成一个样本,则共可抽取 N^n 个样本。例如,总体有 A、B、C、D 这 4 个单位,要从中以重复抽样的方法抽取 2 个单位构成样本。先从 4 个单位中取 1 个,共有 4 种取法,结果登记后再放回,然后再从相同的 4 个中取 1 个,也有 4 种取法,前后取两个构成一个样本,全部可能抽取的样本数目为 $4 \times 4 = 16$ 个,它们是 AA、AB、AC、AD、BA、BB、BC、BD、CA、CB、CC、CD、DA、DB、DC 和 DD。

用不重复抽样的方法,抽取同样多个单位的样本,则全部可能抽取的样本数目为 $N(N-1)(N-2) \cdots (N-n+1)$ 个。例如,从 A、B、C、D 这 4 个单位中,用不重复抽样的方法从中抽取两个单位构成样本。先从 4 个单位中取 1 个,共有 4 种取法;第二次再从留下的 3 个单位中取 1 个,共有 3 种取法;前后两个构成一个样本,全部可能抽取的样本数目为 $4 \times 3 = 12$ 个,它们是 AB、AC、AD、BA、BC、BD、CA、CB、CD、DA、DB、DC。

因此,在相同样本容量的条件下,重复抽样的样本个数总是大于不重复抽样的样本个数。

6.2 抽样误差

6.2.1 抽样误差的概念

抽样误差（sampling error）是指由于随机抽样的偶然因素使各单位的结构不足以代表总体各单位的结构，而引起抽样指标和全及指标之间的绝对离差，如抽样平均数与总体平均数的绝对离差、抽样成数与总体成数之间的绝对离差等。例如，从 1 000 名学生中抽取 100 人进行查体，其平均体重为 55kg，但其他 900 人，有的可能体重恰好是 55kg，但绝大部分的体重与样本的平均体重 55kg 有差别。样本指标与总体指标之间存在绝对离差，这就是抽样误差。抽样调查中误差有以下两个来源：

一个是登记性误差，即在调查过程中，由于主客观原因而引起的误差；另一个是代表性误差，即样本各单位的结构情况不足以代表总体特征而引起的误差。引起代表性误差发生的两种情况：第一，非随机的代表性误差；第二，随机性误差。在第 2 章统计调查中我们已经说过，在抽样调查中，由于被抽选的样本各种各样，被抽中的样本内部各单位标志的构成比例和总体特征有差异，就会出现或大或小的偶然性误差。这种误差理论上或实际上都是不可避免的，只能做到尽量减少这种误差。抽样误差主要是指这种随机性误差，它可以精确计算，并加以控制。

影响抽样误差大小的因素主要有：

（1）样本单位数。在其他条件不变的情况下，样本单位数越大，抽样误差越小；样本单位数越小，抽样误差就越大。这是因为，随着样本单位数的扩大，样本结构就越能反映总体的结构，样本指标就越能代表总体相应的数量特征。当样本扩大到等于全及总体时，则抽样调查就变为全面调查，代表性问题不再存在，因而也就没有抽样误差可言了。

（2）总体内各单位被研究标志的变异程度。在其他条件不变的情况下，总体被研究标志的变异程度越大，抽样误差也越大；总体被研究标志的变异程度越小，抽样误差也越小。这是因为，总体标志变异程度小，表示总体各单位标志值之间的差异小，则样本指标与总体指标之间的差异也可能小。如果总体变异程度为零，即总体内各单位的标志值都相等，样本指标就完全等于总体指标，此时抽样误差也会为零。

（3）抽样方法。重复抽样和不重复抽样是两种不同的抽样方法，抽样方法不同，抽样误差的大小也不同。一般来说，不重复抽样的抽样误差要小于重复抽样的抽样误差。

（4）抽样组织形式。采用不同的抽样组织形式，也会有不同的抽样误差，因为不同的抽样组织形式所抽出的样本对于总体的代表性不同。例如，通常情况下，按照机械抽样和类型抽样方式组织抽样调查，由于经过了排队和分类，可以缩小标志变异程度，因而抽取相同数目的调查单位，其抽样误差会比简单随机抽样方式的抽样误差小。

6.2.2 抽样平均误差

抽样平均误差是反映抽样误差一般水平的指标。通常用抽样平均数的标准差或抽样成数的标准差来作为衡量其抽样误差一般水平的尺度。按照标准差的一般意义,抽样平均数(或成数)的标准差是按抽样平均数(或成数)与其平均数的离差平方和计算的,但由于抽样平均数的平均数等于总体平均数,而抽样成数的平均数等于总体成数,则抽样指标的标准差恰好反映了抽样指标和总体指标的平均离差程度。

$$\mu_{\bar{x}} = \sqrt{\frac{\sum(\bar{x} - \bar{X})^2}{M}} \quad (6-15)$$

$$\mu_p = \sqrt{\frac{\sum(p - P)^2}{M}} \quad (6-16)$$

式中: $\mu_{\bar{x}}$ ——抽样平均数的平均误差;

μ_p ——抽样成数的平均误差;

M ——全部可能的样本数目。

这些公式表明了抽样平均误差的关系。但是由于总体平均数和总体成数并不知道,而且也无法计算全部样本的抽样指标值,所以按上述公式计算抽样平均误差在实际上是不可能的。在实用中可以通过其他方法加以推算。下面分别就抽样平均数和抽样成数的抽样平均误差的计算问题加以讨论。

1. 抽样平均数的平均误差

抽样平均数的平均误差分重复抽样和不重复抽样两种情况。

(1) 重复抽样的抽样平均数的平均误差计算公式如下:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} \quad (6-17)$$

式中: $\mu_{\bar{x}}$ ——抽样平均数的平均误差;

σ^2 ——全及总体方差;

σ ——全及总体标准差(即均方差);

n ——抽样数目。

但是, σ 是全及总体标准差,这是不知道的。如果知道,就无须进行抽样调查了。在这种情况下可用样本的方差 S^2 来代替,其计算公式如下:

$$S^2 = \frac{\sum(x - \bar{x})^2}{n} \quad (6-18)$$

这是简单式方差的计算,如果是加权式,则按下式计算:

$$S^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f} \quad (6-19)$$

(2) 不重复抽样的抽样平均误差计算公式如下:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)} \quad (6-20)$$

式中: N ——全及总体单位数。

当全及总体单位数 N 很大时, $N-1$ 接近于 N , 即可用 N 代替, 则公式 (6-20) 可简化如下:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N} \right)} \quad (6-21)$$

由此可见, 不重复抽样的抽样平均误差公式与重复抽样的抽样平均误差公式的差别, 只在于不重复抽样的抽样平均误差公式的根号下, 乘了一个修正系数 $\left(1 - \frac{n}{N} \right)$ 。因为 $\frac{n}{N}$ 总是一个正小数, 用 1 减去 $\frac{n}{N}$, 其值总是小于 1。而 $\frac{\sigma^2}{n}$ 乘上一个小于 1 的修正系数 $\left(1 - \frac{n}{N} \right)$, 其值总比 $\frac{\sigma^2}{n}$ 的值小, 即 $\frac{\sigma^2}{n} > \frac{\sigma^2}{n} \left(1 - \frac{n}{N} \right)$ 。

因此, 不重复抽样所得到的抽样平均误差比重复抽样所求得的抽样平均误差小。即

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} > \mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N} \right)} \quad (6-22)$$

【实例 6.5】 从某学院 2015 级的 2 000 名学生中, 按简单随机抽样方式抽取 40 名学生, 对公共理论课的考试成绩进行检查, 其平均成绩为 78.75 分, 标准差为 12.13 分。试根据重复抽样和不重复抽样的方法分别计算抽样平均误差。

(1) 按重复抽样计算抽样平均误差:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{12.13}{\sqrt{40}} = 1.92 \text{ (分)}$$

(2) 按不重复抽样计算抽样平均误差:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N} \right)} = \sqrt{\frac{12.13^2}{40} \times \left(1 - \frac{40}{2000} \right)} = 1.90 \text{ (分)}$$

由上述计算结果可以看出, 不重复抽样的抽样平均误差比重复抽样的抽样平均误差小。

在统计工作的实践中, 一般采用不重复抽样的方法进行抽样调查。但其抽样平均误差的计算公式中带有修正系数 $\left(1 - \frac{n}{N} \right)$, 计算起来较为麻烦, 为了计算简便, 一般采用重复抽样的抽样平均误差公式, 这样计算并不影响抽样平均误差的准确性。

2. 抽样成数的抽样平均误差

(1) 重复抽样的抽样成数的平均误差计算公式如下:

$$\mu_p = \sqrt{\frac{P(1-P)}{n}} \quad (6-23)$$

式中: μ_p ——抽样成数的抽样平均误差;

P ——成数;

n ——抽样单位数。

(2) 不重复抽样的抽样平均误差计算, 是在重复抽样平均误差的公式中, 乘上一个修正系数 $\left(1 - \frac{n}{N}\right)$, 即

$$\mu_p = \sqrt{\frac{P(1-P)}{n} \left(1 - \frac{n}{N}\right)} \quad (6-24)$$

式中: N ——全及总体单位数。

公式(6-23)和公式(6-24)中的 P 是总体成数, 这是不知道的; 而 $P(1-P)$ 是总体方差, 它等于 σ^2 , 所以 $P(1-P) = \sigma^2$ 也是不知道的。在这种情况下, 可用样本的成数 P 和样本方差 $P(1-P)$ 来代替, 或者用过去类似调查的成数和成数方差来代替。成数方差的最大值为:

$$P(1-P) = 0.5 \times (1 - 0.5) = 0.5 \times 0.5 = 0.25$$

这就是说, 总体内部两种情况各占 50% 时, 成数方差的最大值为 0.25。

【实例 6.6】 从 40 000 件产品中, 随机抽取 200 件进行检验, 结果有 10 件不合格, 求合格率的抽样平均误差。

$$P = \frac{200 - 10}{200} = 95\%$$

$$\sigma^2 = P(1-P) = 95\% \times (1 - 95\%) = 4.75\%$$

(1) 按重复抽样计算抽样平均误差:

$$\mu_p = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{0.0475}{200}} = 1.54\%$$

(2) 按不重复抽样计算抽样平均误差:

$$\mu_p = \sqrt{\frac{P(1-P)}{n} \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{0.0475}{200} \times \left(1 - \frac{200}{40000}\right)} = 1.53\%$$

以上计算结果表明, 采用不重复抽样的方法算出的成数抽样平均误差比采用重复抽样的方法算出的成数抽样平均误差小。

6.2.3 抽样极限误差

抽样极限误差又称抽样允许误差,是指样本指标与总体指标之间产生抽样误差被允许的最大可能范围,它是根据所研究对象的变异程度和分析任务的要求来确定的可允许的误差范围,凡是在这个范围内的数字都算有效,统计上把这种可允许的误差范围称为抽样极限误差。根据样本指标估计总体指标必然会产生误差,但误差不能太大,否则会使抽样估计失去意义;但误差也不是愈小愈好,因为这样就要增加样本单位数,必然会提高很多费用,所以要确定一个误差范围。

抽样极限误差通常用样本指标可允许变动的上限或下限与总体指标的绝对离差表示。设 $\Delta_{\bar{x}}$ 和 Δ_p 分别表示抽样平均数和抽样成数的抽样极限误差,则有:

$$\Delta_{\bar{x}} = |\bar{x} - \bar{X}| \quad (6-25)$$

$$\Delta_p = |p - P| \quad (6-26)$$

式(6-25)和式(6-26)可变换为下列不等式:

$$\bar{x} - \Delta_{\bar{x}} \leq \bar{X} \leq \bar{x} + \Delta_{\bar{x}} \quad (6-27)$$

$$p - \Delta_p \leq P \leq p + \Delta_p \quad (6-28)$$

式(6-27)和式(6-28)表明总体平均数 \bar{X} 以抽样平均数 \bar{x} 为中心,在 $\bar{x} \pm \Delta_{\bar{x}}$ 的范围内变动。总体成数 P 也是如此,这个变动范围称为估计区间。

6.2.4 抽样极限误差的概率度

抽样极限误差与抽样平均误差不同,它是人为规定的,可大可小。而抽样平均误差是计算出来的,在一定的抽样方式中是唯一的。

基于概率估计的要求,抽样极限误差通常需要以抽样平均误差 $\mu_{\bar{x}}$ 或 μ_p 为标准单位来衡量。把极限误差 $\Delta_{\bar{x}}$ 或 Δ_p 分别除以 $\mu_{\bar{x}}$ 或 μ_p ,得出相对数 t ,表示误差范围为抽样平均误差的 t 倍。 t 是测量估计可靠程度的一个参数,称为抽样误差的概率度,用公式表示为:

$$t = \frac{\Delta_{\bar{x}}}{\mu_{\bar{x}}} = \frac{|\bar{x} - \bar{X}|}{\mu_{\bar{x}}} \quad (6-29)$$

$$t = \frac{\Delta_p}{\mu_p} = \frac{|p - P|}{\mu_p} \quad (6-30)$$

抽样极限误差也可以表示为抽样平均误差的若干倍,其倍数即是概率度 t 。用公式表示为:

$$\Delta_{\bar{x}} = |\bar{x} - \bar{X}| = t \square \mu_{\bar{x}} \quad (6-31)$$

$$\Delta_p = |p - P| = t \square \mu_p \quad (6-32)$$

上述关系式表明，抽样极限误差 Δ 可以用 t 倍的抽样平均误差 μ 表示。在抽样平均误差为一定的条件下，概率度 t 的数值越大，则极限误差 Δ 越大，用样本指标估计总体指标的精确程度也就越低，但估计的可靠程度就越高；反之，若 t 的数值越小，则 Δ 越小，抽样估计的精确程度就越高，但估计的可靠程度就越低。

数理统计能够证明，可靠程度与概率度 t 密切联系，它是 t 的函数，表示为 $F(t)$ ，抽样平均数的分布服从正态分布。从《正态分布概率表》（见附录 B）中可以看出， $F(t)$ 随着 t 的增大而增大，即 t 增大了，可靠性也增大了，但抽样极限误差也扩大了，估计的准确性也降低了。常用概率度与概率保证度（可靠程度）表如表 6.2 所示。

表 6.2 常用概率度与概率保证度（可靠程度）表

概率度 t	1.00	2.00	3.00	1.64	1.96	2.58
概率 $F(t)$ (%)	68.27	95.45	99.73	90.00	95.00	99.00

6.3 抽样推断的方法

6.3.1 抽样估计

抽样估计是指利用实际调查计算的样本指标值来估计相应的总体指标的数值。由于总体指标是表明总体数量特征参数，所以也称为参数估计。总体参数估计有点估计和区间估计两种。

1. 点估计

点估计（point estimation）的基本特点是，根据总体指标的结构形式设计样本指标作为总体参数的估计量，并以样本指标的实际值作为相应总体参数的估计值（如以样本平均数的实际值作为相应总体平均数的估计值，以样本成数的实际值作为相应总体成数的估计值等）。我们之所以作这样的考虑，是基于我们对所研究的总体指标的具体指标值虽然不知道，但是对于它的指标结构形式都是清楚的。

例如，抽选 6 000 名大学生，对其实际月消费支出进行调查，结果表明这 6 000 名大学生的月平均消费支出为 326 元，我们就推断说，全体大学生的月消费支出为 326 元。

对总体参数作估计的时候，总是希望估计是合理或优良的。那么优良估计的标准是什么呢？所谓优良估计总是从总体上来评价的，其标准有三个方面。

（1）无偏性（unbiasedness）。即以抽样指标估计总体指标要求抽样指标值的平均数等于被估计的总体指标值本身。就是说，虽然每一次的抽样指标值和总体指标值之间都可能存在误差，但在多次反复的估计中，各个抽样指标值的平均数应该等于所估计的总体指标值本身，即对抽样指标的估计平均说来是没有偏误的。

从 6.2 节已经知道，抽样平均数的平均数等于总体平均数，抽样成数的平均数等于总体成数，即

$$E(\bar{x}) = \bar{X} \quad (6-33)$$

$$E(p) = P \quad (6-34)$$

这说明以抽样平均数作为总体平均数的估计量,以抽样成数作为总体成数的估计量,是符合无偏性原则的。

(2) 一致性 (consistency)。当以抽样指标估计总体指标要求当样本的单位数充分大时,抽样指标也充分地靠近总体指标。就是说,随着样本单位数 n 的无限增加,抽样指标和未知的总体指标之差的绝对值小于任意小的数,它的可能性也趋过于必然性,即实际上是几乎肯定的。

我们知道,抽样平均数和抽样成数的抽样平均误差和样本单位数的平方根成反比例变化,样本单位数愈多则平均误差便愈小,当样本单位数接近于总体单位数时,平均误差也就接近于零。也就是说,抽样平均数和抽样成数作为总体平均数和总体成数的估计量,是符合一致性原则的。

(3) 有效性 (efficiency)。以抽样指标估计总体指标要求作为优良估计量的方差应该比其他估计量的方差小。例如,用抽样平均数或总体某一变量值来估计总体平均数,虽然两者都是无偏的,而且在每一次估计中,两种估计量和总体平均数都可能有离差,但样本平均数更靠近于总体平均数的周围,平均说来其离差比较小。所以对比说来,抽样平均数是更有效的估计量。

总体参数点估计方法的优点是简便、易行、原理直观,常为实际工作者所采用。但也有不足之处,这种估计没有表明抽样估计的误差,更没有指出误差在一定范围内的概率保证程度有多大。要解决这个问题,就必须采用总体参数的区间估计方法。

2. 区间估计

(1) 区间估计的概念和要素。区间估计 (interval estimation) 是在一定的概率把握程度下,根据样本指标和抽样极限误差去估计总体指标所在可能范围的方法。这个范围通常用一个最低限和一个最高限构成的区间来表示,并以一定的概率把握程度保证总体指标的估计值在这两个数值构成的区间之内。在总体指标的区间估计公式中,有两个要素:一个是置信区间,另一个是置信概率。

置信区间 (confidence interval) 又称为估计区间,是指由低限到高限两个数值所构成的可能范围,由样本指标和极限误差组成。置信区间是一个随机区间,这是由其组成要素的随机性决定的。置信区间越小,允许的抽样误差范围也越小,则估计的准确性越高;反之,估计的准确性越低。

置信概率 (confidence probability) 是指区间估计的概率保证程度,也称为置信度。它是表明样本指标与总体指标的误差不超过一定范围的概率有多大,在样本指标的基础上,运用极限误差进行估计,总是要和一定的概率保证程度联系在一起的。因为我们采用的是随机抽样,抽样误差是一个随机变量,则不能保证误差不超过一定范围是一个必然事件,而只能给予一定程度的概率保证。所以,在进行抽样估计时,不仅要考虑极限误差的大小,而且还要考虑被估计

的数值落在这个范围的概率的大小。前者是估计的准确性问题，后者是估计的可靠性问题，两者紧密联系且不可分割。

(2) 总体平均数和总体成数的区间估计。

① 总体平均数的估计。总体平均数的估计就是用抽样平均数来估计总体平均数。

【实例 6.7】 某学校进行一次英语测验，为了解学生的考试情况，随机抽选部分学生进行调查，所得资料如表 6.3 所示。

表 6.3 学生考试情况

考试成绩 (分)	60 以下	60~70	70~80	80~90	90 以上
学生人数 (人)	10	20	22	40	8

要求：按重复抽样方法以 95.45% 的概率估计该校学生英语考试的平均成绩的范围。

解：由已知条件计算抽样平均数、标准差：

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{7\,660}{100} = 76 \text{ (分)}$$

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2 f}{\sum f}} = \sqrt{\frac{12\,944}{100}} = 11.38 \text{ (分)}$$

抽样平均误差为：

$$\mu_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{11.38}{\sqrt{100}} = 1.14 \text{ (分)}$$

抽样极限误差为：

$$\Delta_{\bar{x}} = t\mu_{\bar{x}} = 2 \times 1.14 = 2.28 \text{ (分)}$$

$$\text{下限} = \bar{x} - \Delta_{\bar{x}} = 74.32 \text{ (分)}$$

$$\text{上限} = \bar{x} + \Delta_{\bar{x}} = 78.88 \text{ (分)}$$

所以，在 95.45% 的概率保证程度下，该校学生英语考试的平均成绩的范围为 74.32 分 ~ 78.88 分。

② 总体成数的估计。总体成数的估计就是用抽样成数来估计总体成数。

【实例 6.8】 利用【实例 6.7】给出的资料，以同样的概率估计该校学生成绩在 80 分以上的学生所占的比重范围。

解：根据资料可得：

$$\text{样本合格率为：} P = \frac{48}{100} = 48\%$$

$$\text{抽样平均误差为: } \mu_p = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{0.48 \times (1-0.48)}{100}} = 4.996\%$$

$$\text{抽样极限误差为: } \Delta_p = t\mu_p = 2 \times 4.996\% = 9.992\%$$

$$\text{下限} = p - \Delta_p = 38.01\%$$

$$\text{上限} = P + \Delta_p = 57.99\%$$

所以，在 95.45% 的概率保证程度下，该校学生成绩在 80 分以上的学生所占的比重范围为 38.01%~57.99%。

6.3.2 样本容量的确定

1. 确定必要样本容量的意义

样本单位数是决定抽样误差的大小的直接因素，因此，在组织抽样调查时，必须事先确定样本单位数。样本单位数越多，样本的代表性越大，抽样误差越小，抽样估计就越可靠。但样本容量过多会增加不必要的人力、物力和费用开支，造成浪费。样本容量减少，会使抽样误差增大，达不到所需要的准确程度。可见，合理确定必要的样本单位数，以保证取得较好的抽样估计效果，是组织抽样调查的重要问题。

确定必要样本容量的基本原则是，在保证预期的抽样估计可靠程度的要求下，抽取的样本单位数不宜过多。

2. 影响样本容量的因素

为了确定必要的样本容量，我们首先分析决定样本容量的各种主要因素。

(1) 总体各单位标志变异程度。总体标志变异程度大，要求样本容量大；反之，总体标志变异程度小，样本容量小。

(2) 抽样方法。在其他条件相同的情况下，重复抽样要比不重复抽样多抽取一些样本单位。

(3) 抽样组织形式。一般来说，类型抽样和等距抽样的样本容量要小于简单随机抽样的样本容量。

(4) 极限误差的大小。如果允许误差大，样本容量就小；如果允许误差小，样本容量就大。

(5) 抽样估计的可靠程度，即概率 $F(t)$ 的大小。如果估计的可靠程度要求越高，即 $F(t)$ 越大，样本容量就越多；如果估计的可靠程度要求越低，即 $F(t)$ 越小，样本容量就越少。

以上影响样本容量的因素，可以从样本容量的计算公式加以印证。

3. 必要样本容量的计算公式

(1) 平均数的必要样本容量的计算公式。

在重复抽样的条件下为：

$$n = \frac{t^2 \sigma^2}{\Delta_x^2} \quad (6-35)$$

在不重复抽样的条件下为:

$$n = \frac{N t^2 \sigma^2}{N \Delta_x^2 + t^2 \sigma^2} \quad (6-36)$$

(2) 成数的必要样本容量的计算公式。

在重复抽样的条件下为:

$$n = \frac{t^2 P(1-P)}{\Delta_p^2} \quad (6-37)$$

在不重复抽样的条件下为:

$$n = \frac{N t^2 P(1-P)}{N \Delta_p^2 + t^2 P(1-P)} \quad (6-38)$$

【实例 6.9】 对某罐头厂生产的罐头质量进行抽样调查, 抽样极限误差为 5%, 概率为 0.954 5, 并知过去进行同样抽样调查, 其不合格率为 10%, 试求必要的样本容量。

根据题意可知: $t=2$, $P=90\%$, $\Delta_p=5\%$ 。

在重复抽样的条件下:

$$n = \frac{t^2 P(1-P)}{\Delta_p^2} = \frac{2^2 \times 90\% \times (1-90\%)}{(5\%)^2} = 144 \text{ (盒)}$$

4. 计算必要样本容量应注意的问题

(1) 上述公式计算的样本容量是最低的, 也是最必要的样本容量。

(2) 用上述公式计算样本容量时, 一般情况下, 总体方差是未知的, 处理方法与“抽样平均误差计算公式”相同。

(3) 如果进行一次抽样调查, 同时对总体平均数和成数进行估计, 此时就会计算出两个样本容量, 并且二者通常情况下又是不相等的。那么, 为了同时满足两个估计的要求, 应该在两个样本容量中选择较大的一个进行抽样。

(4) 上述公式计算的样本容量不一定是整数。如果带有小数, 不允许采取四舍五入的办法化为整数, 而是用比这个小数稍大的邻近整数代替。

6.3.3 抽样的组织形式

抽样推断是根据事先规定的要求而设计的抽样调查组织, 并以所获得的这一部分实际资料为基础, 进行推理演算做出结论。如何科学地组织抽样调查, 保证随机抽样条件的实现, 并合

理有效地取得各项实际数据，成为抽样中一个至关重要的问题。本节主要介绍几种常用的抽样组织形式，简单随机抽样、分层抽样、等距抽样、整群抽样和多阶段抽样等。

1. 简单随机抽样

简单随机抽样（simple random sample）又称纯随机抽样，它是按随机原则直接从总体 N 个单位中抽取 n 个单位作为样本。无论是重复抽样还是不重复抽样，都要保证每个单位在抽选中都有相等的中选机会。简单随机抽样是最基本、最简单的抽样组织形式，它适用于均匀总体，即具有某种特征的单位均匀地分布于总体的各个部分，使总体的各部分都是同分步的。通常抽取样本可采用直接抽选法（直接从调查对象中随机抽选），也可在抽样前对总体各单位加以编号，然后用抽签的方式或根据《随机数表》（见附录 C）来抽选必要的单位数。

直接抽取法是从调查总体中直接随机抽取样本进行调查。这种方法适合对集中于某个较小空间的总体进行抽样，如对存放于仓库的同类产品直接随机抽出若干产品为样本进行质量检查。

抽签法是将总体各单位编上序号并将号码写在外形相同纸片上掺和均匀后，再从中随机抽取。被抽中的号码所代表的单位，就是随机样本，直到抽够预先规定的样本数目为止。如在某城市某街道所管辖的 10 000 户居民中，抽 200 户居民面向居民对某种商品的需求量进行调查，就可以做 10 000 张纸片，写上 1~10 000 号，从中随机抽取 200 张，即得到被抽中的居民为样本。一般在随机抽样中，用“ N ”代表总体单位数，用“ n ”代表样本单位数。显然这个问题中 $N=10\,000$ 个， $n=200$ 个。

随机数表法是先把总体各单位编号，根据编号的最大数（即总体单位数位数确定使用随机数表中若干列或若干行数字），从任意行或任意列的第一个数字起，可以向任何方向去数，遇到属于总体单位编号范围内的号码就定为样本单位，直到抽够预定的样本单位数为止。随机数表是由 0~9 组成的表，这 10 个数字排列完全是随机的。较大的随机数表有美国兰德公司 1955 年编制出版的 100 万数字表和肯德尔与史密斯在 1938 年编制出版的 10 万数字表（附录 C 是美国兰德公司 1955 年编制出版的随机数表中的 1 000 个数字）。随机数表法比抽签简化，它免去了对每个总体单位做签的过程，这在总体单位数 N 较大时，尤其显示出它的便捷。

纯随机抽样是随机抽样最基本的方法，也是其他随机抽样法的基础。它完全符合随机原则，且应用简单易行，但这种方法的不足在于：在总体很大的情况下使用时，编号工作量繁重；当总体单位差异程度较大时，必须使样本容量充分大才能保证样本推断总体的可靠程度和准确程度；所抽取的样本在总体中的分布或是过于集中，或是过于分散，很不均匀，给实际调查带来困难，在实践应用中有不便之处。

2. 分层抽样

分层抽样（stratified sampling）又称类型抽样，它是先对总体各单位按主要标志加以分组，然后再从各组中按随机原则抽选一定单位构成样本。

例如，调查在校大学生每月生活费支出情况，为保证样本单位能容纳各种类型的学生，可将大学生按性别分为男、女两组，然后在每组中随机抽选若干样本单位进行调查；或按年级分组来抽选样本单位，这样就能够保证样本有较充分的代表性。

将总体分成若干组后, 在各组中抽选样本单位有以下三种方法。

第一, 按各组的标志变异程度来确定各组的应抽单位数。对于标志变异大的组宜多抽一些单位进行调查; 对标志变异小的组宜少抽一些单位进行调查。这种分配方法称为样本单位最佳分配。

第二, 按调查者主观意识任意确定各组应抽选单位数, 这种分配方法称为随意分配。

第三, 按统一的比例确定各组应抽选的单位数, 这种方法称为比例分配。

本节主要讨论按比例分配样本单位的方法。比例分配使下列条件存在:

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \cdots = \frac{n_k}{N_k} = \frac{n}{N} \quad (6-39)$$

式中: k ——总体分成的组数;

n_1, n_2, \cdots, n_k ——各组应抽选的单位数;

N_1, N_2, \cdots, N_k ——各组中的总体单位数。

并且 $\sum n_i = n$, $\sum N_i = N$; 因而使 $n_i = N_i \cdot \frac{n}{N} = n \cdot \frac{N_i}{N}$ 。

如果按比例分配样本单位, 可以得到如下类型抽样误差公式。

(1) 在重复抽样条件下:

变量总体:
$$\mu_{\bar{x}} = \sqrt{\frac{\sigma_i^2}{n}} \quad (6-40)$$

式中: $\overline{\sigma_i^2} = \frac{\sum \sigma_i^2 n_i}{n}$ 是各组方差的加权平均数, 习惯上称为平均组内方差。

属性总体:
$$\mu_p = \sqrt{\frac{P_i(1-P_i)}{n}} \quad (6-41)$$

式中: $\overline{P_i(1-P_i)} = \frac{\sum P_i(1-P_i)n_i}{n}$ 是各组比率的方差加权平均数, 是属性总体的平均组内方差。

(2) 在不重复抽样条件下:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma_i^2}{n} \left(1 - \frac{n}{N}\right)} \quad (6-42)$$

$$\mu_p = \sqrt{\frac{P_i(1-P_i)}{n} \left(1 - \frac{n}{N}\right)} \quad (6-43)$$

式中的总体参数 σ_i^2 和 P_i 是未知的, 习惯上用样本值 S_i^2 和 P_i 代替。

【实例 6.10】 对某乡全部 4 000 亩土地按类型抽样了解该乡平均亩产和全乡小麦总产量，其中有平原地 3 000 亩，丘陵地 1 000 亩，按 3%比例抽取样本单位，调查结果如表 6.4 所示。要求在 95%的概率保证下对全乡小麦平均产量和总产量进行估计。

表 6.4 某乡小麦产量抽样资料

土地类型	全部面积 N_i (亩)	抽样面积 n_i (亩)	抽样单位产量 x_i (斤)	各组抽样平均数 \bar{x}_i (斤)	样本方差 σ_i^2
平原	3 000	9	420、420、450 460、465、470 480、490、520	464	900
丘陵	1 000	3	300、320、340	320	256
合计	4 000	12	5 135	—	—

注：1 (市) 亩=0.066 7 公顷；1 (市) 斤=0.5 千克。

解：抽取平原地亩数： $n_i = N_i \times \frac{n}{N} = 3\,000 \times 3\% = 9$ (亩)

抽取丘陵地亩数： $n_i = N_i \times \frac{n}{N} = 1\,000 \times 3\% = 3$ (亩)

抽样平均数： $\bar{x} = \frac{\sum \bar{x}_i n_i}{n} = \frac{464 \times 9 + 256 \times 3}{12} = 428$ (斤)

抽样平均误差： $\mu_{\bar{x}} = \sqrt{\frac{\sigma_i^2}{n} (1 - \frac{n}{N})} = \sqrt{\frac{739}{12} \times (1 - \frac{12}{4\,000})} = 7.84$ (斤)

式中： $\sigma_i^2 = \frac{\sum \sigma_i^2 n_i}{n} = \frac{900 \times 9 + 256 \times 3}{12} = 739$ (斤)

全乡平均亩产区间为：

$$\bar{X}: \bar{x} \pm t\mu_x = 428 \pm 1.96 \times 7.84$$

即在 $412.63 \leq \bar{X} \leq 443.37$ (斤) 范围内。

全乡小麦总产量区间为：

$$412.63 \times 4\,000 \leq \sum X \leq 443.37 \times 4\,000$$

即在 $1\,650\,520 \leq \sum X \leq 1\,773\,480$ (斤) 范围内。

另外，在计算必要抽样单位数目中，用 $\overline{\sigma^2}$ 取代简单随机抽样中的 σ^2 ，即可得到类型抽样 n 的数目。如在重复条件下的变量总体，它的样本单位应为：

$$n = \frac{t^2 \overline{\sigma_i^2}}{\Delta_{\bar{x}}^2} \tag{6-44}$$

有关区间估计的方法和步骤，与简单随机抽样一致。

3. 等距抽样

等距抽样 (systematic sampling) 也称机械抽样或系统抽样, 它先按某一标志对总体各单位进行排序, 然后依一定顺序和间隔来抽取样本单位的一种抽样组织。由于这种抽样是在各单位大小按顺序排列的基础上, 再按某种规则依一定间隔取样, 所以可以保证所取得的样本单位比较均匀地分布在总体的各个部分, 有较高的代表性。

与类型抽样一样, 在抽样前选择了某一标志划分总体, 这是它们的一致之处。不一致之处主要在于:

第一, 类型抽样选择标志是与研究目的有重要关系的标志, 而机械抽样既能选择与研究目的有某种关系的标志, 也能选择与研究目的无关的标志。例如, 调查员工工资水平, 如用类型抽样, 就必须选择与工资水平这一调查内容有密切联系的“工龄”标志将总体分组, 分成“10年以下”、“10~20年”、“20年以上”等组; 如用机械抽样, 可选择与调查内容无关的“姓氏笔画”或“身高”等标志将所有工人按照顺序排列编号。

第二, 类型抽样选择标志目的是为了将总体单位划分为有限个组 (组数较少), 在每组中抽取 n_i 个样本单位; 而机械抽样选择标志的目的是为了将总体单位顺序排列, 划成 n 个相同单位部分或组 (组数较多), 在每组中仅抽取一个样本单位。

第三, 类型抽样选择分组标志可以是品质标志和数量标志, 如调查居民收入, 可按照“职业”品质标志分组, 也可按照“工龄”数量标志分组; 但是机械抽样只选数量标志排列, 如上述选“姓氏笔画”、“身高”等数量标志将员工顺序排列编号。

机械抽样根据总体单位数 (N) 和需抽选的样本单位数 (n), 可将总体单位划成 n 个相等部分, 每部分都包括 $K = \frac{N}{n}$ 个单位, 然后在第一个部分中抽选一个单位, 依 K 个间隔依次在各部分都抽取一个单位组成样本。

机械抽样的重要工作是选取第一个样本单位。一般说来, 确定方法主要有以下三种。

第一种, 随机取样。从第一部分包括的 K 个单位中随机确定任意 i 号单位为第一个样本单位, 然后依次在第二部分中取 $(i+K)$ 号为第二个样本单位, 第三部分取 $(i+2K)$ 号……第 n 部分取 $[i+(n-1)K]$ 号, 直至取 n 个单位为止。由上可见, 当第一个样本单位 i 号被确定后, 其他各样本单位随之也被确定下来。

【实例 6.11】 某街道有 10 000 名职工, 现需统计他们的收入情况, 拟抽 100 名职工进行调查。按照与研究目的有关的标志“工龄”, 将 10 000 名职工顺序排列, 将每名职工按 00001、00002……10000 编上相应号码, 然后将 10 000 名职工分成 100 个相等部分, 每部分 $K = \frac{N}{n} = \frac{10\,000}{100} = 100$ 名职工, 第一部分包括 00001~00100 号, 第二部分为 00101~00200 号……

按随机原则在第一部分抽取 00014 号为第一个样本单位, 那么根据依次加 $K=100$ 个间隔抽取样本单位的要求, 其号码是: 00114、00214、00314……09814、09914 号, 即 100 个单位组成样本。

第二种，中位取样。中位取样与随机取样不同之处是第一个样本单位不是随机抽选，而是以第一部分的中间位置（或靠近中间位置）的号码为第一个样本单位号码，即以 $\frac{K+1}{2}$ （或 $\frac{K}{2}$ ）号码为第一个中选单位。依上例用中位取样，依次抽取的号码是：00050、00150、00250……09950。在城市住户调查中，按照有关标志排列进行中位取样的方法比随机取样方法代表性好，因此得到比较广泛的应用。

第三种，对称取样。对称取样是在中位取样的基础上改进的，它是在每部分的两侧按对称位置抽选样本单位。如果随机地在第一部分抽取第一个样本单位 i 号，那么在第二部分抽取第二个样本单位为 $(2K-i)$ 号，第三个样本单位则在 $(2K+i)$ 号，以后依次为 $(4K-i)$ 、 $(4K+i)$ 、 $(6K-i)$ 、 $(6K+i)$ ……即在单数部分抽取 K 的偶倍数加 i 号，在双数部分抽取 K 的偶倍数减 i 号。这样所抽选的结果使各部分的样本单位标志值大小具有相互抵消而趋于平均数的趋向，使样本代表性增加，抽样误差减小。

仍依上例，如采用对称取样，从第一部分取第 00030 号（ $i=30$ ）为第一个样本单位，则在第二部分取第 00170 号（ $2K-i=200-30$ ）为第二个样本单位，依次抽取的号码是：00230、00370……09830、09970。这种对称样本在按有关标志排列下能够取得比类型抽样更小的误差。因此，国家统计局要求农产量调查、城乡住户调查尽可能使用此方法。

机械抽样的抽样误差直接计算很困难，可用下列方法处理：第一，从理论上说，按无关标志排列的机械抽样，用简单随机抽样误差公式计算；按有关标志排列的机械抽样，用类型抽样误差公式计算。但在实践中，都用简单随机抽样误差公式计算。

第二，机械抽样都采用不重复方法抽选样本单位，因而都用不重复抽样误差公式计算。

机械抽样在总体单位标志数值差异程度比较大时运用，其优越性更能得到充分发挥。

4. 整群抽样

整群抽样（cluster sampling）又称集团抽样，它是将总体单位划分成 R 群（组），每群（组）包括 M_i 个单位，然后从 R 群中随机地抽选若干群（ r 群），对每个抽中群的 M_i 个单位进行全面调查的一种组织方式。

例如，进行城市住户调查，将全市住户按居委会为基本单位分成 R 群，然后将各居委会按某个标志排队编号，或等距或随机地抽选 r 个居委会，对抽中的居委会的全部住户进行调查。

整群抽样的优点在于：组织工作简单、搜集资料方便容易、调查费用较少，但是抽样误差一般都比较大。即使要得到与简单随机抽样相同的精确度，整群抽样也要调查大得多的样本单位。

整群抽样是以群为抽取基本单位，抽一次得到的包括 M_i 个单位为一群。每群大小可以相同（如批量生产的产品质量调查，每批数目大致相同），也可以不同（如上例每个居委会住户数目不同）。通常我们调查的各群 M_i 相同时，用简单平均法计算样本指标；各群 M_i 不相同时，用加权平均法计算样本指标。

整群抽样都采用不重复抽样的方法, 因此计算抽样误差公式都用不重复抽样公式。

变量总体:

$$\mu_{\bar{x}} = \sqrt{\frac{\delta^2}{r} \left(\frac{R-r}{R-1} \right)} \quad (6-45)$$

式中: δ^2 ——组间方差。

当每群大小相同时用简单式:

$$\delta^2 = \frac{\sum (x_i - \bar{x})^2}{r} \quad (6-46)$$

式中: \bar{x}_i ——各群平均数, $\bar{x}_i = \frac{\sum x_i}{M_i}$;

\bar{x} ——样本平均数, $\bar{x} = \frac{\sum \bar{x}_i}{r}$ 。

当每群大小不同时用加权式 (公式从略)。

属性总体:

$$\mu_p = \sqrt{\frac{\delta^2}{r} \left(\frac{R-r}{R-1} \right)} \quad (6-47)$$

当每群大小相同时用简单式:

$$\delta^2 = \frac{\sum (p_i - p)^2}{r} \quad (6-48)$$

式中: p_i ——各群抽样成数, $p_i = \frac{M_{i1}}{M_i}$;

p ——样本成数, $p = \frac{\sum p}{r}$ 。

当每群大小不同时用加权式 (略)。

必要抽样群数 r 的确定公式类同简单随机抽样中有关公式。只不过现在确定的是 r , 并用组内方差 δ^2 取代 σ^2 。

$$r = \frac{Rt^2\delta^2}{R\Delta^2 + t^2\delta^2} \quad (6-49)$$

【实例 6.12】 某产品合计生产 3 000 批, 每批产量大致相同, 抽样检查产品合格品率, 要求其允许误差不超过 2%, 若根据过去该产品质量资料, 其各批群间方差为 3%, 问在 95.45% ($t=2$) 可靠程度下需抽多少批产品为宜?

$$r = \frac{Rt^2\delta^2}{R\Delta^2 + t^2\delta^2} = \frac{3\,000 \times 2^2 \times 0.03}{3\,000 \times 0.02^2 + 2^2 \times 0.03} \approx 273 \text{ (批)}$$

即要在全部产品中抽 273 批产品检查。

5. 多阶段抽样

多阶段抽样 (multi-stage sampling) 是在组织抽样时, 不是从总体中一次直接抽取样本单位, 而是把抽样过程分成几个过渡阶段进行, 到最后才具体抽取样本单位。它是先抽大单位, 再在大单位中抽小单位, 在小单位中抽更小的单位, 直到抽出最终能取得推断总体的基本单位为止。如进行全省企业职工家庭状况调查, 先在全省中抽市, 在抽中的城市的各部中抽选企业, 最后在抽中的企业中抽选职工, 进行职工家庭情况调查。

一般在总体很大, 单位分布面积广, 倘若从总体中采取一次直接抽取足够多的样本很困难时, 可采用多阶段抽样。

多阶段抽样的具体步骤是: 首先, 将调查总体各单位按一定标志分成若干集体, 作为抽样的第一级单位; 然后将第一级单位再分成若干小集体, 作为抽样的第二级单位; 依次类推, 按研究问题的需要和现象本身的特点, 分出第三级单位、第四级单位等。其次, 依照随机原则, 先在第一级单位中抽出若干单位作为第一级单位样本; 然后在第一级被抽中的单位中再抽出第二级单位样本; 依次类推, 还可抽出第三级单位样本、第四级单位样本, 由此形成了两阶段随机抽样、三阶段随机抽样或四阶段随机抽样等。

在多阶段抽样中, 前几阶段的抽样, 都类似整群抽样, 每一阶段抽样都会存在抽样误差。为提高抽样指标的代表性, 各阶段抽取群数的安排和抽样方式, 都应注意样本单位的均匀分布。

首先, 适当多抽第一阶段的群数, 使样本单位在总体中得到均匀分布。但是, 需要注意的是, 样本过于分散则需要更多的人力和经费。

其次, 根据方差的大小, 来考虑各阶段抽取群数的多少。对于群间方差大的阶段, 应当适当多抽一些群; 反之, 则可少抽一些群。

最后, 各阶段抽样时, 可以根据条件, 将各种抽样组织方式灵活运用, 而且尽可能利用现成资料。

以两阶段抽样为例, 首先将总体划分为 R 组, 而每组包含 M_i 个单位。第一步从 R 组中随机抽取 r 组; 第二步, 再从中选的 r 组中分别随机抽取 m_i 个单位, 构成一个样本, 这种抽样就是两阶段抽样。其中, 总体单位数 $N=M_1+M_2+\cdots+M_R$, 各组的单位数 M_i 可以是相等的, 也可以是不等的。样本单位数 $n=m_1+m_2+\cdots+m_r$, 各组抽取的样本单位可以是相等的, 也可以是不等的。为简化起见, 假定 R 组中各组的单位数相等, 都为 M , 则有 $N=RM$, 而且从各组抽取的单位数也相等, 都为 m , 则有 $n=rm$ 。

两阶段抽样和类型抽样、整群抽样的共同点是它们都须先对总体加以分组, 然后再抽取单位, 但它们之间却有明显的差别。类型抽样是从全部的分组中每组各抽取单位, 它和两阶段抽样的区别在于第一阶段取了全部的组, 而两阶段抽样在第一阶段只是随机地抽取部分的组。整

群抽样是从全部的分组中随机抽取部分的组，然后对中选的全部单位进行调查。它和两阶段抽样的区别在于第二阶段抽取了中选组的全部单位，而两阶段抽样在第二阶段只是在中选组中随机地抽取部分单位。所以，两阶段抽样在组织技术上是整群抽样和类型抽样的综合。

6.4 参数假设检验

6.4.1 假设检验的基本概念

假设检验(hypothesis testing)是利用样本的实际资料来检验事先对总体某些数量特征所作的假设是否可信的一种统计分析方法。它和参数估计一样，都是利用样本资料对总体特征进行某种推断。但二者推断角度不同，参数估计是根据样本指标以一定的把握程度估计总体参数取值范围，更准确地说，是对总体参数进行了区间估计；而假设检验则先对总体参数值提出一个假设，然后利用样本信息，以一定的概率水平去判断这个假设是否成立。下面的一些例子，都是假设检验问题。

【实例 6.13】 某食品公司生产一种罐头，按标准每罐净重为 227 克，根据以往生产经验，罐头重量的标准差为 3 克。现随机抽查该公司产品 100 罐，测得平均净重为 228 克，判断这批罐头是否符合标准？

【实例 6.14】 某轮胎制造商在广告中声称，该公司生产的汽车轮胎在正常行驶条件下平均寿命高于 28 000 公里。检测部门随机挑选了 40 个轮胎进行测试，结果显示平均行驶里程为 28 200 公里，样本标准差为 1 000 公里，问该制造商广告是否属实？

【实例 6.15】 某批发商欲从厂家购进一批灯泡，根据合同规定，灯泡的平均使用寿命不能低于 1 000 小时。从产品中随机抽取 100 只灯泡，测得平均寿命为 960 小时，标准差为 200 小时，请问批发商是否应该购买这批灯泡？

从直观上看，上面例子中样本指标数值与总体指标数值的要求有一些差异，但这种差异可能是由于抽样的随机性带来的，事实上也许并没有显著差异。究竟是否存在显著差异，可以先对总体设立一个假设，然后检验这个假设是否成立，这就是一个假设检验问题。由此可见，假设检验是对我们所关心的，却又是未知的总体参数先做出假设，然后抽取样本，利用样本提供的信息对假设的正确性进行判断的过程。它是进行经济管理和决策的重要工具。

6.4.2 假设检验的步骤

一个完整的假设检验过程，通常包括以下四个步骤：

第一，提出原假设和备择假设；

第二，选取和计算检验统计量；

第三，根据显著性水平查临界值；

第四, 进行比较并做出决策。

下面, 我们分别对每一个步骤的内容加以分析和说明。

1. 提出原假设和备择假设

在统计学中, 把需要通过样本去推断其正确与否的命题称为原假设, 用 H_0 表示。而与原假设相对立的假设便是备择假设, 用 H_1 表示。

在对总体均值进行假设检验时, 用 \bar{x}_0 表示某特定的数值, 所提出的假设共有三种情况。

$$\begin{array}{lll} \text{第一种: } H_0: \bar{X} = \bar{x}_0 & \text{第二种: } H_0: \bar{X} \geq \bar{x}_0 & \text{第三种: } H_0: \bar{X} \leq \bar{x}_0 \\ H_1: \bar{X} \neq \bar{x}_0 & H_1: \bar{X} < \bar{x}_0 & H_1: \bar{X} > \bar{x}_0 \end{array}$$

请注意, 等号总是在原假设上。

第一种情况的假设检验, 我们称为双侧检验 (two-tailed test)。第二、三种情况的假设检验, 我们称为单侧检验 (one-tailed test)。其中第二种称为左单侧检验, 第三种称为右单侧检验。

当我们面临实际问题进行检验时, 如何提出假设? 可以按照下面三步来做。首先, 将所要检验的问题用符号表述。在例 6.13 中要检验的问题是“罐头是否符合标准”, 表述为 $\bar{X} = \bar{x}_0$; 在例 6.14 中要检验广告是否属实, 即汽车轮胎平均寿命是否高于 28 000 公里, 表述为 $\bar{X} > \bar{x}_0$; 例 6.15 所关心的是“批发商是否应该购买这批灯泡”, 而据合同, 实际上检验的是灯泡的平均使用寿命是否不低于 1 000 小时, 表述为 $\bar{X} \geq \bar{x}_0$ 。

其次, 写出刚才所表述式子的余集。例 6.13 其余集为 $\bar{X} \neq \bar{x}_0$, 例 6.14 的余集为 $\bar{X} \leq \bar{x}_0$, 而例 6.15 的余集为 $\bar{X} < \bar{x}_0$ 。

最后, 根据“等号总是在原假设上”, 找出原假设和备择假设。例 6.13 中 $H_0: \bar{X} = \bar{x}_0$, $H_1: \bar{X} \neq \bar{x}_0$, 实为双侧检验; 例 6.14 中 $H_0: \bar{X} \leq \bar{x}_0$, $H_1: \bar{X} > \bar{x}_0$, 为右单侧检验; 而例 6.15 中原假设为 $H_0: \bar{X} \geq \bar{x}_0$, 备择假设为 $H_1: \bar{X} < \bar{x}_0$, 是左单侧检验。

2. 选取和计算检验统计量

在参数假设检验中, 如同在参数估计中一样, 要借助于样本统计量进行统计推断。在具体问题中如何选统计量, 需要考虑一些因素, 如用于检验的样本是大样本还是小样本, 总体方差已知还是未知等。

对总体均值的假设检验, 在大样本 ($n \geq 30$) 情况下, 用 Z 统计量, 即 $Z = \frac{\bar{x} - \bar{X}_0}{\sigma / \sqrt{n}}$; 在小样本 ($n < 30$) 情况下, 用 t 统计量, 即 $t = \frac{\bar{x} - \bar{X}_0}{s / \sqrt{n}}$ 。不同情况下选用不同统计量, 这些在后面具体分析。

3. 根据显著性水平查临界值

由于假设检验是利用样本的信息对总体进行推断，就有可能犯错误。例如，原假设正确，而我们却把它当成错误的加以拒绝。犯这种错误的概率用 α 表示， α 称为假设检验中的显著水平（level of significance）。显著水平是指，当原假设正确时，人们却把它拒绝的概率。这个概率是由具体问题确定的，通常取 $\alpha=0.05$ 或 $\alpha=0.01$ 。这表明，当作出接受原假设的决定时，其正确的可能性（概率）为 95%或 99%。

α 为显著水平，其所对应的概率度称为显著性水平 α 的临界值。例如， $\alpha=0.05$ 时，在正态分布情况下可以直接利用概率表查找临界值作为判断依据。若为双侧检验， $-Z_{0.025}=-1.96$ 、 $Z_{0.025}=1.96$ 为两个临界值，当 Z 统计量值在 $[-1.96, 1.96]$ 内时，接受原假设。若为右单侧检验，则 $\alpha=0.05$ ，查表时，其临界值相当于双侧检验中 $\alpha=0.1$ 所对应的右临界值点，应为 $Z_{0.05}=1.645$ ，当 Z 统计量值 ≤ 1.645 时，接受原假设。若为左单侧检验，其临界值为 $-Z_{0.05}=-1.645$ ，当 Z 统计量值 ≥ -1.645 时，接受原假设。

4. 进行比较并做出决策

用计算出的检验统计量的值与临界值相比较，可以做出接受或拒绝原假设的结论，进而可以对所检验的实际问题进行决策。

6.4.3 假设检验中的两类错误

假设检验可能犯两种类型的错误。第一类错误（type I error）指的是 H_0 客观上真实但被检验所拒绝，这种错误也称为弃真错误，犯这种错误的概率就是显著水平 α ；第二类错误（type II error）指的是 H_0 客观上不真实但被检验所接受，这种错误也称为纳伪错误，犯这种错误的概率我们用 β （读作“贝塔”）来表示。表 6.5 列出不同实际情况，不同决策对应的 4 种可能结果的概率。

表 6.5 两类错误

	H_0 真实	H_0 不真实
接受 H	正确决策 $(1-\alpha)$	第二类错误 (β)
拒绝 H_0	第一类错误 (α)	正确决策 $(1-\beta)$

人们自然希望犯两种错误的概率越小越好，但这是不可能的，因为两者是矛盾的。为使 α 变小就要尽量接受 H_0 ，其结果必然导致 β 加大。对此，假设检验中一般是保证 α 小到一定标准前提下，再来考虑尽量缩小 β ，即优先考虑控制“弃真”风险。常用的标准是 $\alpha=0.05$ 、 0.025 、 0.01 ，在经济领域中常用 $\alpha=0.05$ 。

6.4.4 总体均值和总体成数检验

1. 大样本情况下总体均值检验

在大样本（ $n \geq 30$ ）情况下，总体均值的假设检验可以应用正态分布检验法。

例如, 对【实例 6.13】以 $\alpha=0.05$ 的显著水平进行假设检验。

解: 第一步提出原假设和备择假设。 $H_0: \bar{X}=227$, $H_1: \bar{X} \neq 227$ 。

第二步选取和计算检验统计量。 $Z = \frac{\bar{x} - \bar{X}_0}{\sigma / \sqrt{n}}$, 代入数值, 计算得 $Z=3.33$ 。

第三步根据显著性水平查临界值。由 $\alpha=0.05$, 得临界值 $-Z_{0.025}=-1.96$, $Z_{0.025}=1.96$ 。

第四步进行比较并做出决策。 $\because Z=3.33 > Z_{0.025}=1.96$

\therefore 拒绝 H_0

即这批罐头不符合标准。

而对【实例 6.14】进行假设检验 ($\alpha=0.05$) 时应采用单侧检验。

解: 第一步提出假设。这需要确定检验的方向, 题中制造商称轮胎寿命高于 28 000 公里, 表述为 $\bar{X} > 28\,000$, 其余集为 $\bar{X} \leq 28\,000$, 由于等号在原假设上, 故 $\bar{X} \leq 28\,000$ 为原假设 H_0 , $\bar{X} > 28\,000$ 为备择假设 H_1 , 此为右单侧检验。

第二步计算统计量。由于总体标准差 σ 未知, 用样本标准差 S 替代。 Z 统计量为 $\frac{\bar{x} - \bar{X}_0}{s / \sqrt{n}}$, 代入数值得 $Z=1.26$ 。

第三步查临界值。因为是右单侧检验, $\alpha=0.05$ 时其临界值 $Z_{0.05}=1.645$ 。

第四步比较并做出决策。 $\because Z=1.26 < Z_{0.05}=1.645$

\therefore 接受 H_0

\therefore 轮胎平均使用寿命不高于 28 000 公里

\therefore 制造商广告不属实。

我们再对【实例 6.15】以 $\alpha=0.05$ 的显著水平进行假设检验。

解: 第一步提出假设。 $H_0: \bar{X} \geq 1\,000$, $H_1: \bar{X} < 1\,000$ 。

第二步计算检验统计量。采用 Z 统计量 $\frac{\bar{x} - \bar{X}_0}{s / \sqrt{n}}$, 代入数值, 计算得 $Z=-2$ 。

第三步查临界值。由 $\alpha=0.05$, 左单侧检验得临界值 $-Z_{0.05}=-1.645$ 。

第四步比较并做出决策。 $\because Z=-2 < -Z_{0.05}=-1.645$

\therefore 拒绝 H_0

即批发商不应当购买这批灯泡。

2. 小样本情况下总体均值检验

在小样本 ($n < 30$) 情况下, 总体均值的假设检验可以应用 t 分布检验法。此时, 检验统计量为 t 统计量, $t = \frac{\bar{x} - \bar{X}_0}{s / \sqrt{n}}$, 它服从自由度为 $n-1$ 的 t 分布。

【实例 6.16】 某食品公司生产袋装食品, 采用自动打包机打包。每包标准净重应为 1 000 克, 每天开工后需要检验一次打包机工作是否正常。现从产品中随机抽取 9 包, 实测每包净重如下: 987、993、1 012、1 005、997、983、1 021、995、1 005, 给定显著性水平 $\alpha=0.05$, 问该日打包机工作是否正常?

解: 由于本题关心的是每包净重是否为 1 000 克, 因此是双侧检验问题, 又因为是小样本资料, 故采用 t 分布检验法。

第一步提出假设。 $H_0: \bar{X}=1\,000$, $H_1: \bar{X} \neq 1\,000$ 。

第二步计算检验统计量。根据 $t = \frac{\bar{x} - \bar{X}_0}{s / \sqrt{n}}$, 样本平均数 $\bar{x}=999.8$, 样本标准差 $S=12.12$, 代入数值, 计算得 $t=-0.049\,5$ 。

第三步查临界值。 $\alpha=0.05$, 由于是双侧检验, 自由度 $=n-1=8$, 查 t 分布表, 得临界值 $-t_{0.025}(8)=-2.306$, $t_{0.025}(8)=2.306$ 。

第四步比较并做出决策。 $\because -t_{0.025}(8)=-2.306 < t=-0.049\,5 < t_{0.025}(8)=2.306$

\therefore 接受 H_0

\therefore 该日打包机工作正常。

3. 总体成数检验

在统计实际工作中, 经常需要检验总体单位中含有某种特征的单位数所占的比例是否为某个假设值 P_0 , 如一批产品中的合格品率、在一次民意测验中表示赞同的比例等。此时通常采用大样本资料, 因而可以用 Z 统计量进行检验。方法及步骤与总体均值检验一样, 只是在统计量公式中需要用到成数的标准差。其统计量公式为:

$$Z = \frac{p - P_0}{\sqrt{P_0(1 - P_0) / n}}$$

【实例 6.17】 一家杂志声称其读者中至少有 25% 是在校大学生。从该杂志读者中抽取 200 名读者进行调查, 发现其中 42 人是在校大学生。请问该杂志的论断是否正确? ($\alpha=0.05$)

解: 第一步提出假设。 $H_0: p \geq 25\%$, $H_1: p < 25\%$ 。

第二步计算检验统计量。 $p=n_1/n=42/200=21\%$, 根据公式:

$$Z = \frac{p - P_0}{\sqrt{P_0(1 - P_0) / n}}$$

代入数值，计算得到 $Z=-1.31$ 。

第三步查临界值。由 $\alpha=0.05$ ，左单侧检验得临界值 $-Z_{0.05}=-1.645$ 。

第四步比较并做出决策。 $\because Z=-1.31 > -Z_{0.05}=-1.645$

\therefore 接受 H_0

即该杂志的论断正确。

6.5 Excel在抽样推断中的应用

6.5.1 利用Excel进行区间估计

在 Excel 中，我们可以利用其提供的一些统计函数进行样本推断总体的分析。一般主要用到的函数有平均数函数 AVERAGE、样本标准差函数 STDEV、 t 分布函数 TINV，利用这些函数构造一个用于推断统计的工作表。操作方法借助于【实例 6.18】进行说明。

【实例 6.18】 从某高校 2011 级经济管理系 400 名学生中随机抽取 40 名学生，获取统计学课程成绩如下，进行样本推断总体的分析，计算结果如图 6.1 所示。

69 61 75 60 80 72 82 86 90 78
62 87 71 77 79 87 85 97 85 63
94 78 65 85 96 88 89 82 87 78
95 83 56 81 58 75 78 74 67 75

	A	B	C	D	E	F	G
1	样本数据						
2	56		样本个数	40			
3	58		样本均值	78.25			
4	60		样本标准差	10.85510633			
5	61		标准误差	1.716343012			
6	62		置信水平	95%			
7	63		自由度	39			
8	65		t值	2.022688932			
9	67		置信区间半径	3.471628013			
10	69		置信区间上限	81.72162801			
11	71		置信区间下限	74.77837199			
12	72						

图 6.1 推断统计结果

解：具体的操作步骤如下。

第一步：输入样本数据。

第二步：输入相关变量的名称。本例中为 C 列的内容，相对应的 D 列为变量值所在单元格。

第三步：计算要求的量的数值。

(1) 计算样本个数。单击 D2 单元格，输入 “=COUNT(A:A)”，回车后在单元格内显示出样本数据的个数，即 A 列中的数据个数。

(2) 计算样本均值。单击 D3 单元格，输入 “=AVERAGE(A:A)”，回车后在单元格内显示出样本均值。

(3) 计算样本标准差。单击 D4 单元格，输入 “=STDEV(A:A)”，回车后在单元格内显示出样本标准差。

(4) 计算样本标准误差。单击 D5 单元格，输入 “=D4/SQRT (D2)”，回车后在单元格内显示出样本标准误差。样本标准误差=样本标准差/SQRT (样本个数)。

(5) 输入置信度。单击 D6 单元格，输入 “=95%”，回车后在单元格内显示出置信度 95%。

(6) 计算自由度。单击 D7 单元格，输入 “=D2-1”，回车后在单元格内显示出自由度 39 (自由度=样本个数-1)。

(7) 计算 t 值。单击 D8 单元格，输入 “=TINV(1-D6, D7)”，回车后在单元格内显示出 t 值 (t 值=TINV(1-置信度, 自由度))。

(8) 计算置信区间半径。单击 D9 单元格，输入 “=D8*D5”，回车后在单元格内显示出置信区间半径的值 (置信区间半径=t 值×标准误差)。

(9) 计算置信区间的上限。单击 D10 单元格，输入 “=D3+D9”，回车后在单元格内显示出置信区间的上限的数值 (置信区间的上限=样本均值+置信区间半径)。

(10) 计算置信区间的下限。单击 D11 单元格，输入 “=D3-D9”，回车后在单元格内显示出置信区间的下限的数值 (置信区间的下限=样本均值-置信区间半径)。

6.5.2 利用Excel进行假设检验

【实例 6.19】 外贸公司出口一批食品，规定标准为每袋净重 150 克。检验人员随机抽取了 30 包实测每包净重，结果如表 6.6 所示。

表 6.6 30 包食品净重									单位：克
148.0	148.5	149.0	148.3	148.7	150.0	151.9	150.3	151.8	149.9
150.8	151.1	149.4	150.2	150.7	151.5	149.2	150.6	150.2	150.1
151.8	150.8	150.4	150.2	151.2	151.4	149.5	150.1	151.3	151.1

给定显著性水平 $\alpha=0.05$ ，请问该批食品是否符合标准？请利用 Excel 进行假设检验。

解：由题意可知，本检验为双侧检验，原假设 $H_0: \bar{X}=150$ ；备择假设 $H_1: \bar{X} \neq 150$ 。

由 $\alpha=0.05$ ，得临界值 $-Z_{0.025}=-1.96$ ， $Z_{0.025}=1.96$ 。

具体的操作步骤如下。

第一步：输入样本数据。

第二步：在“插入函数”对话框中选择“统计”选项，找到 STDEV 后单击“确定”按钮，打开对话框。

第三步：在 Number1 中选择样本数据区域“A2:A31”，然后就可直接得到样本标准差 $S=1.072$ 。

第四步：同样在“插入函数”对话框中选择“统计”选项，找到 ZTEST 后单击“确定”按钮，打开对话框。

第五步：在 Array 中选择样本数据区域“A2:A31”；在 X 文本框中输入均值 150；在 Sigma 中输入样本标准差 1.072，就可直接得到 Z 统计量， $Z=0.0865$ 。

第六步：比较后得结论。

因为 $-1.96 < Z=0.0865 < 1.96$ ，所以接受 H_0 ，即该批食品符合标准。

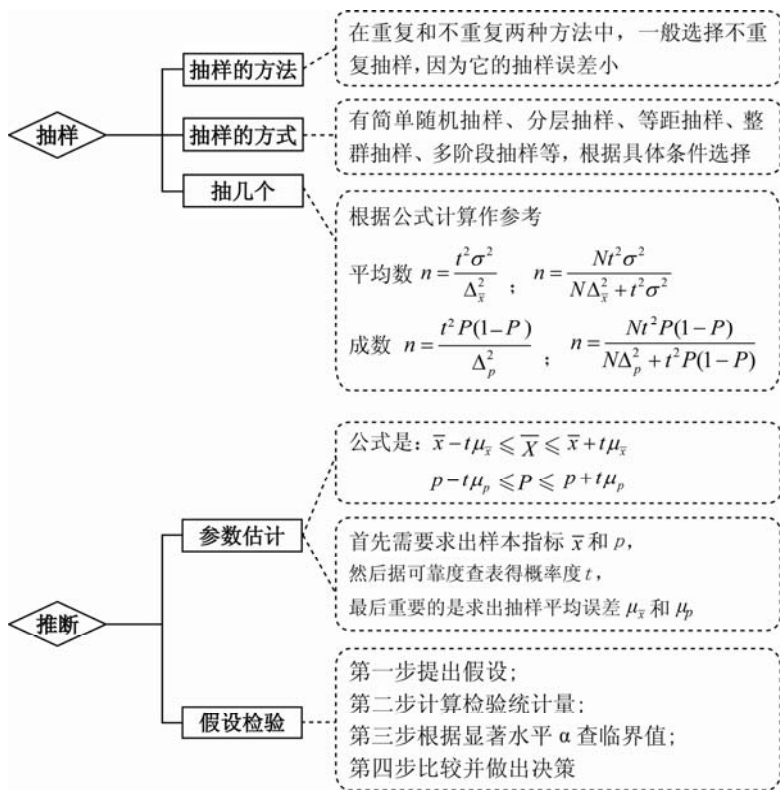


统计术语

总体	population	样本	sample
参数	parameter	统计量	statistic
重置	with replacement	不重置	without replacement
抽样推断	sampling inference	抽样调查	sampling survey
抽样误差	sampling error	点估计	point estimation
区间估计	interval estimation	简单随机抽样	simple random sample
等距抽样	systematic sampling	分层抽样	stratified sampling
整群抽样	cluster sampling	多阶段抽样	multi-stage sampling
置信区间	confidence interval	置信概率	confidence probability
假设检验	hypothesis test	原假设	null hypothesis
备择假设	alternative hypothesis	第一类错误	type I error
第二类错误	type II error	显著性水平	level of significance
单侧检验	one-tailed test	双侧检验	two-tailed test



重点知识梳理



习题与实践训练

一、判断题

1. 抽样推断中, 作为推断对象的总体和作为观察对象的样本都是确定的, 唯一的。 ()
2. 样本容量是指从一个总体中可能抽取的样本个数。 ()
3. 样本成数是指在样本中具有某种性质的单位数在全部单位数中所占的比重。 ()
4. 抽样极限误差总是大于抽样平均误差。 ()
5. 在总体方差一定的条件下, 样本单位数越多, 则抽样平均误差越大。 ()
6. 如果要检验“某产品的平均重量是否显著大于 100 克”, 应采用右单侧检验。 ()
7. 在其他条件不变的情况下, 提高抽样估计的可靠程度, 可以提高抽样估计的精确度。 ()
8. 样本单位数的多少与总体各单位标志值的变异程度成反比, 与抽样极限误差范围的大小成正比。 ()
9. 在假设检验中, 原假设错误但却被接受, 这种错误称为第二类错误。 ()
10. 在简单重复随机抽样中, 如果抽样极限误差扩大为原来的 2 倍, 其他条件不变, 则样本单位数只需要原来的四分之一。 ()

二、单项选择题

- 在抽样推断中, 必须遵循的原则是 ()。
 - 随意原则
 - 随机原则
 - 可比原则
 - 对等原则
- 能够事先加以计算和控制的误差是 ()。
 - 抽样误差
 - 登记性误差
 - 系统性误差
 - 测量误差
- 当我们所要检验的是样本所取自总体的参数值是偏高或偏低某个特定值时, 应选择 ()。
 - 双侧检验
 - 单侧检验
 - 左单侧检验
 - 右单侧检验
- 抽样推断中的概率保证程度表达了区间估计的 ()。
 - 显著性
 - 准确性
 - 可靠性
 - 规律性
- 抽样平均误差是 ()。
 - 全部样本指标的平均数
 - 全部样本指标的平均差
 - 全部样本指标的标准差
 - 全部样本指标的标志变异系数
- 在同等条件下, 重复抽样与不重复抽样相比较, 其抽样平均误差 ()。
 - 前者小于后者
 - 前者大于后者
 - 两者相等
 - 无法确定哪一个大
- 抽样极限误差和抽样平均误差之间的关系为 ()。
 - 极限误差可以大于、等于或小于平均误差
 - 极限误差一定大于平均误差
 - 极限误差一定小于平均误差
 - 极限误差一定等于平均误差
- 当抽样平均误差为 0.8, 极限误差为 2 时, 则概率度为 ()。
 - $t=2$
 - $t=2.5$
 - $t=3$
 - $t=3.5$
- 在总体内部情况复杂, 单位数较多, 且各单位之间的变异程度较大时, 宜采用 () 进行调查。
 - 整群抽样
 - 分类抽样
 - 纯随机抽样
 - 等距抽样
- 在一定抽样平均误差的条件下 ()。
 - 缩小极限误差, 可以提高推断的可靠程度
 - 缩小极限误差, 推断的可靠程度不变
 - 扩大极限误差, 可以提高推断的可靠程度
 - 扩大极限误差, 可以降低推断的可靠程度

三、多项选择题

- 抽样调查的特点是 ()。
 - 按随机原则抽取样本
 - 按随意原则抽取样本
 - 由部分推断总体
 - 可以事先计算并控制抽样误差
 - 缺乏科学性和可靠性
- 按组织方式不同, 抽样调查有 ()。

- A. 纯随机抽样
 - B. 等距抽样
 - C. 类型抽样
 - D. 整群抽样
 - E. 重复抽样和不重复抽样
3. 抽样调查中的抽样误差是 ()。
- A. 不可避免的
 - B. 可以事先计算并加以控制
 - C. 抽样估计值与总体参数值之差
 - D. 可以避免的
 - E. 受总体标志变动程度的影响
4. 影响抽样误差的主要因素有 ()。
- A. 抽样数目的多少
 - B. 总体标志变异程度的大小
 - C. 不同的组织方式
 - D. 抽样周期的长短
 - E. 不同的抽样方法
5. 要提高抽样推断的精确度, 可采用的方法有 ()。
- A. 增加样本数目
 - B. 减少样本数目
 - C. 缩小总体被研究标志的变异程度
 - D. 改善抽样的组织方式
 - E. 改善抽样的方法
6. 影响抽样数目的主要因素有 ()。
- A. 总体被研究标志的变异程度大小
 - B. 抽样的组织方式
 - C. 对推断精确度的要求
 - D. 对推断把握程度的要求
 - E. 抽取调查单位的方法
7. 在简单重复随机抽样条件下, 欲使误差范围缩小 $\frac{1}{2}$, 其他要求保持不变, 则样本容量必须 ()。
- A. 比原来增加 2 倍
 - B. 比原来增加 3 倍
 - C. 是原来的 4 倍
 - D. 比原来减少 2 倍
 - E. 是原来的 3 倍
8. 在抽样推断中 ()。
- A. 抽样指标的数值不是唯一的
 - B. 总体指标是一个随机变量
 - C. 可能抽取许多个样本
 - D. 统计量是样本变量的函数
 - E. 全及指标又称为统计量
9. 简单随机抽样 ()。
- A. 适用于总体各单位呈均匀分布的总体
 - B. 适用于总体各单位标志变异较大的总体

- C. 在抽样之前要求对总体各单位加以编号
 - D. 最符合随机原则
 - E. 是各种抽样组织形式中最基本、最简单的一种形式
10. 在其他条件不变的情况下, 抽样极限误差与置信度的关系是 ()。
- A. 极限误差越小, 置信度越大
 - B. 极限误差越小, 置信度越小
 - C. 极限误差越大, 置信度越大
 - D. 成正比关系
 - E. 成反比关系

四、填空题

1. 抽样推断是在_____的基础上, 利用样本资料计算样本指标, 并据以推算_____特征的一种统计分析方法。
2. 在重复抽样的条件下, 抽样平均误差与_____成反比, 与_____成正比。
3. 简单随机抽样在抽取样本单位时有重复抽样和_____两种不同的方法。
4. 对被研究标志变动较大的总体进行抽样推断时, 宜采用_____组织方式调查。
5. _____反映了样本指标与总体指标之间的抽样误差的可能范围。
6. _____反映了样本指标与总体指标之间的平均误差程度。
7. 计算抽样平均误差, 若未知总体平均数方差时, 可用_____来代替。若未知总体成数方差时, 可以用_____代替。
8. 假设检验中的假设一般包括两部分, 即_____和_____。
9. 样本单位数达到或超过_____称为大样本。
10. 抽样调查过程中可能发生的误差分为_____和_____两类。

五、应用能力训练题

1. 从外贸进口的一批商品 10 000 件中随机抽取 100 件检验其质量, 发现有 10 件不合格。试按重复与不重复抽样方法分别计算产品合格率抽样平均误差。
2. 在一项新家电产品的市场调查, 我们随机选取 400 位顾客为样本, 询问他们是否喜欢此产品, 其中 72.1% 的顾客表示喜欢该产品, 试在 95% 的概率保证下, 估计该产品在顾客中的占有率。
3. 某广告公司想估计某类商店去年所花的平均广告费有多少, 经验表明, 总体方差为 1 800 000, 置信度为 95%, 则:
 - (1) 要使估计值处在总体平均数附近 500 元的范围内, 应抽取多少个商店?
 - (2) 若抽取 56 个商店, 那么估计值处在总体平均数附近多少元的范围内?
4. 从一批出口产品中按不重复随机抽样方法抽选 200 件, 检测出废品 8 件, 又知道抽样数是该批产品总数的 1/20, 当概率为 0.954 5 时, 可否认为这批产品的废品率不超过 5%?
5. 一位美国学者对美国政治历史感兴趣, 他主持了一项美国总统身高研究, 想知道美国总统身高与美国普通男人身高相比是否一致。他搜集了 20 世纪美国 17 位总统的身高资料, 得出如下数据: 17 位总统平均身高 180.638 8cm, 标准差 8.584 6cm, 美国成年男人平均身高 176cm (数据来源: Triola, M.F.(2005)Elementary Statistics)。则在 0.05 显著性水平下, 是否可以推断

美国总统的身高与普通男人身高有显著差异？

6. 从某年级学生中按简单随机抽样方式抽取 100 名学生，对会计学科的考试成绩进行检查，及格的有 82 人，试以 95.45% 概率保证程度推断全年级学生的及格率区间范围。如果其他条件不变，将允许误差缩小一半，应抽取多少名学生检查？

7. 外贸公司出口一种食品，规定每包规格不低于 150 克，现在用不重复抽样的方法抽取其中的 100 包进行检验，其结果如下。

每包重量（克）	包数（包）
148～149	10
149～150	20
150～151	50
151～152	20
合 计	100

根据以上资料：

(1) 以 99.73% 概率估计这批食品平均每包重量的范围，以便确定平均重量是否达到规格要求。

(2) 以同样的概率估计这批食品合格率范围。

8. 某公司接受调查一种新型洗衣机的销售前景，它拟通过抽样调查，了解居民在收听（看）有关广告节目后对新型洗衣机感兴趣的户数比例。据从前推销其他商品的经验，这个比例不会超过 20%，若确定置信区间宽度为 4 个百分点，取置信水平为 95%，样本容量是多少？

9. 某学校有 2 000 名学生参加英语等级考试，为了解学生的考试情况，用不重复抽样方法抽取部分学生进行调查，所得资料如下。

考试成绩（分）	60 以下	60～70	70～80	80 以上
学生人数（人）	20	20	45	15

试以 95.45% 的概率估计该校学生英语等级考试成绩在 70 分以上的学生所占比重范围。

10. 某居民小区有居民 2 800 户，小区的物业管理者准备采用一项新的管理措施，想了解居民是否赞成，采取抽样方法随机抽取了 140 户，其中有 90 户赞成，50 户反对。请利用 Excel 软件以 95% 把握程度，求出小区中赞成该项改革的户数比例的置信区间。



本章案例

2011 年下半年，《纽约时报》头版报道了“美国全国陷入了对未来的深深忧虑和怀疑中”，作者对美国人的心理进行了探究，整理了美国公众对奥巴马政府的表现、社会财富分配等众多问题的普遍看法。下面，我们就来了解一下 2011 年秋天美国人想要表达的想法。

- 有高达 89% 的美国人不相信政府会做正确的事——美国政府遭遇了有记录以来最严峻的一次信任危机。
- 有 2/3 的美国公众认为，财富应该在美国得到更加公平的分配。

- 有 43% 的美国人他们说大体上认同“占领华尔街”运动所宣扬的观点。此外,还有更多的美国人(46%)认为“占领华尔街”运动中抗议人群的观点“基本上反映了绝大多数美国人的观点”。
- 有 46% 的美国人认可奥巴马作为美国总统的工作表现,同样有 46% 的美国人不可奥巴马的工作表现。
- 仅有 9% 的美国公众认可美国国会的工作。
- 虽然距离下一次的美国总统初选只剩下不足 2 个月的时间,但是,还有将近 80% 的共和党选民觉得“现在就决定支持谁为时尚早”。

在美国总统选举年即将到来之际,这些引人入胜的数据可以为人们提供一些有意义的参考,让读者窥见美国人作为一个整体的所思所想。但是,总是有人忍不住要问:我们是如何知道这些情况的?美国的人口超过 3 亿,为什么我们就能对他们的想法做出如此精确的推断?我们怎么知道这些言之凿凿的判断是否正确?

答案只有四个字:抽样推断。这项民意测验是由《纽约时报》和哥伦比亚广播公司(CBS)共同主导的(连两家彼此竞争的媒体都必须在某个民调项目上通力合作,可见要主导一个方法论上可行和完善的全美国性民调有多么“浪费资金”)。《纽约时报》和 CBS 的这次联合民调动用了专业的民意测验机构,基于电话访问,在 6 天时间里他们通过电话调查了 1 650 名美国成年人,其中有 1 475 名美国成年人声称自己是登记选民。

至于具体是如何抽样的,我们只能进行一个大概的猜测,绝大多数的民意测验采用的都是如下的技术。为了保证接电话的人能够代表美国人口,抽样过程是从概率开始的——相当于从口袋中摸彩球。电脑会随机抽取一个座机电话交换机组(电话交换机是汇集电话线路并完成用户之间通话的设备。在美国,一个电话交换机包含一个区号及电话号码的前 3 位),通过在美国约 6.9 万个家庭交换机组里随机选取与电话人口比例一致的用户样本,就能大体上形成一个具有人口地域代表性样本分布。每组被抽中的交换机由电脑随机加上 4 位数字,以形成一个完整的电话号码,最后出现在被呼叫家庭的名单里。同时,该调查还包括了“手机号码的随机拨打”。

举个简单的例子,假如样本为 500 人,有 53% 将选票投给了美国共和党候选人,45% 投给了美国民主党,还有 2% 投给了第三方的候选人。如果以美国共和党的支持率作为参照,此次抽样的平均误差为

$$\mu_p = \sqrt{\frac{P(1-P)}{n}} = 0.02236$$

为方便起见,约等于 2%。

假如你在一家电视台工作,节目制片人向你咨询:我们能否以这次测试的结果作为宣布共和党获胜的依据?作为懂得抽样推断的你应该这样解释,你愿意为播出内容的错误承担多大的风险?若是把握程度为 68% ($t=1$,即承担 32% 的风险),则美国共和党支持率的置信区间是 (51%,55%),民主党是 (43%,47%)。但这个承担错误播报的风险太高了,若是把握程度为 95% ($t=1.96$,即承担 5% 的风险),则美国共和党支持率的置信区间是 (49%,57%),民主党是 (41%,49%)。然而,新的问题出现了,我们无法推翻两党候选人打成平手(各获得 49% 选票)的可能性。这是一个无法避免的妥协,在没有新数据补充的情况下,如果想要提供民调结果的正确率,就只能降低预测的精度。就在此时,转机来了,第二次民意测验的结果出现在你的办

公桌上。这一次的样本容量为 2 000 人，占比结果为：共和党 52%、民主党 45%、独立党派 3%。你的制片人已经彻底发疯了，因为这一次的民意测验显示两个主要党派之间的差距进一步缩小了，也就是说，在官方结果出来之前对选举进行预测变得难上加难。但此时你（英勇地）指出，这一次的样本数量是上一次数量的 4 倍，因此抽样误差会大大缩小。共和党候选人的新的抽样平均误差是 $\mu_p = \sqrt{\frac{P(1-P)}{n}} = 0.1$ 。假如制片人还愿意接受 95% 的正确率，那么你便可以大声地宣布共和党将赢得选举。因为在此情况下，95% 的置信区间意味着共和党获得了 50%~54% 的选票，民主党获得了 43%~47% 的选票，两个置信区间之间不再有重叠，你可以在电视上恭喜美国共和党候选人了，而且这次预测正确的概率超过 95%。

在这个例子中，你还可以做得更加完美。如果将可靠程度扩大到 99.7%，那么两党的投票情况是：共和党获得的选票为 $52\% \pm 3\%$ ，即 49%~55%；民主党获得的选票为 $45\% \pm 3\%$ ，即 42%~48%。介于两党的结果依然没有重叠，你便放心地在电视上预测共和党的胜利，你和制片人基本上不可能因为误播而被辞退。所以，一定记得要请那次组织 2 000 人民意测验的同事吃饭哦。

（资料来源：《赤裸裸的统计学》【美】Charles Wheelan 著，曹槟译，中信出版社，2013）

时间数列

学习要点

- 时间数列的概念与种类。
- 时间数列各种分析指标的计算方法。
- 长期趋势的测定方法。
- 季节比率的含义与应用。

社会经济现象总是随着时间的推移而不断地发展变化的。如果我们在若干个连续时期内对某一个经济变量进行测算，搜集到“过去”的时间里大量的数据资料，采用科学方法，从中总结出规律性的结论，然后结合“目前”的客观现状，就可以对“未来”做出判断和预测，这将是经济生活中很有意义的事情！实际上，生活中许多值得我们关注的经济变量都是时间数列的数据。例如，国内生产总值（GDP）每季度测算一次；消费者价格指数和失业率每月测算一次；货币的供应量每周测算一次；而像上海、深圳证券市场上沪深 300 指数（收盘价）则逐日公布。如何将这些有用的信息汇编成科学的时间数列？怎样描绘、分析数列中的指标？时间数列的数据是否具有规律性变动特征，又该如何反映其发展趋势并据此对未来进行预测？本章将简要介绍有关时间数列的理论和方法。

7.1 时间数列的概念与种类

7.1.1 时间数列的概念

所谓时间数列（time-series），就是将某一统计指标的数值按时间的先后顺序排列起来所形成的数列，又称为动态数列。例如，由 2009—2014 年我国国内生产总值所构成的时间数列如表 7.1 所示。

表 7.1 2009—2014 年我国国内生产总值情况

单位：亿元

年 份	2009	2010	2011	2012	2013	2014
国内生产总值	340 903	408 903	484 124	534 123	588 019	636 463

（资料来源：国家统计局）

从表 7.1 中可以看出，时间数列一般是由两个基本要素构成的：一是被研究现象所属的时间，如表 7.1 中的 2009 年、2010 年等；二是现象在各个时间上的统计指标数值，如表 7.1 中的国内生产总值。在时间数列中，指标数值也称为发展水平。

时间数列在统计和经济分析中，有着极为重要的作用。当我们回顾以往工作进行总结时，可以通过对时间数列的编制和分析，描绘社会经济现象变化的过程；当我们立足现在对现状态势进行分析时，可以运用时间数列，将不同地区或国家进行对比分析；当我们展望未来进行经济预测时，可以研究时间数列中现象发展变化的方向、速度、水平、趋势和规律性，为社会经济现象的预测提供依据。

7.1.2 时间数列的种类

时间数列按其统计指标表现的不同形式可分为三种，即总量指标时间数列、相对指标时间数列和平均指标时间数列。其中总量指标时间数列是最基本的时间数列，相对指标时间数列和平均指标时间数列则是在其基础上派生而成的数列。

1. 总量指标时间数列

总量指标时间数列又称为绝对数时间数列，是由一系列同类总量指标按时间顺序排列形成的时间数列，它反映某种社会经济现象在一段时间内达到的绝对水平及其增减变化情况。例如，表 7.1 就反映了我国在 2009—2014 年间各个时期国内生产总值所达到的绝对水平。总量指标时间数列按指标所反映的社会经济现象的时间状况不同，又可分为时期指标数列和时点指标数列两种，简称时期数列和时点数列。

(1) 时期数列及其特点。在总量指标时间数列中，如果指标数值都是反映某种社会经济现象在一定时期内发展过程的总量，则这种时间数列就称为时期数列。表 7.1 所列的数列就是一个时期数列，其中每项指标分别反映 2009—2014 年间各年的国内生产总值的总量。时期数列的特点是：

- ① 时期数列的每一指标数值总是和一定的时期相对应的，数列中各指标值可以累加，相加之后表示现象更长一段时期内的发展总量。
- ② 时期数列中每个指标数值的大小与其所属的时期长短有关，时期越长，指标数值越大；时期越短，指标数值越小。
- ③ 时期数列中各指标数值通常是通过经常性调查及连续不断登记、汇总而取得的。

(2) 时点数列及其特点。在总量指标时间数列中，如果指标数值都是反映某种社会经济现象在某一时刻（时点）上的状况及水平，则这种时间数列就称为时点数列。如表 7.2 所示，表中的数列就是时点数列。

表 7.2 2009—2014 年年末全国总人口数资料 单位：万人

年 份	2009	2010	2011	2012	2013	2014
年末全国总人口数	133 450	134 091	134 735	135 404	136 072	136 782

(资料来源：国家统计局)

与时期数列相对应，时点数列的特点如下：

- ① 时点数列中的每一指标只表明社会经济现象在一定时点上的水平，各指标值不能累加，若相加就产生了重复计算，且无实际意义。
- ② 时点数列中每个指标数值的大小与其时间间隔长短无直接关系。
- ③ 时点数列中各指标数值，是通过一次性调查登记取得的，即时点数列不具有连续统计的特点。

在运用统计资料进行分析时，正确区分时期数列和时点数列具有十分重要的意义，因为根据这两种数列计算平均发展水平时，其计算方法是截然不同的。

2. 相对指标时间数列

相对指标时间数列又称为相对数时间数列，是由一系列相对指标数值按时间顺序排列形成的时间数列，它反映社会经济现象之间数量对比关系的发展变化过程和趋势。由于相对指标一般表现为两个相关的总量指标之比，因此，两个时期指标、两个时点指标或一个时期指标、一个时点指标的对比，都可以形成相对指标时间数列。例如，表 7.3 就是一个由“第三产业增加值”与“国内生产总值”这两个时期指标之比所形成的相对数时间数列。

表 7.3 2009—2014 年我国第三产业增加值占国内生产总值比重资料

年 份	2009	2010	2011	2012	2013	2014
第三产业增加值占国内生产总值比重（%）	42.6	43.0	43.1	44.6	46.1	48.2

（资料来源：国家统计局）

由于相对指标数值的基数不同，因此，相对指标时间数列中的各项数值不能直接相加。

3. 平均指标时间数列

平均指标时间数列又称为平均数时间数列，是由一系列同类平均指标数值按时间顺序排列形成的时间数列，它反映社会经济现象一般水平的变化过程和发展趋势。如表 7.4 所示，就是平均指标时间数列。

表 7.4 “十二五”期间某公司职工年平均工资

单位：元/人

年 份	2011	2012	2013	2014	2015
职工年平均工资	43 944	52 676	59 521	67 560	75 696

平均指标时间数列中的各项指标数值也不能相加，因为加总后的结果不具有实际意义。

总量指标时间数列、相对指标时间数列和平均指标时间数列这三种时间数列，实际上并不是彼此孤立的。基于社会经济现象的复杂性和关联性，在很多情况下，为了准确、全面地分析事物发展变化过程，必须把这三种时间数列结合起来运用。

7.1.3 时间数列的编制原则

编制时间数列的主要目的是，通过对数列中一系列指标数值的研究分析，说明社会经济现

象的发展变化过程及其规律性。因此，保证数列中各项指标数值的可比性，是编制时间数列的基本要求。编制时间数列应遵循以下基本原则。

1. 时期长短应相等

时期数列中各项指标数值的大小与其所属的时期长短有直接关系，因此，在同一个时期数列中，各指标数值所属时期长短要力求相等，否则时期长短不一，很难直接做出判断和比较。当然，对时间可比性的理解不能绝对化，有时为了特殊的研究目的或是信息资料的局限，也可以将时期不等的指标数值编制成时间数列。例如，我国不同时期钢产量的动态数列如表 7.5 所示。

表 7.5 我国几个重要时期钢产量统计表					单位：万吨
年 份	1900—1949	1953—1957	1981—1985	1986—1990	1991—1995
钢产量	776	1 667	20 304	27 372	42 478

从表 7.5 中的数列可以看出，我国钢铁工业在新中国成立以后飞速发展，特别是改革开放以来取得了伟大的成就。

对于时点数列，虽然指标数值的大小与时间间隔长短无直接关系，但也要尽量保持时点间隔的一致，以利于对比分析问题。

2. 总体范围要一致

总体范围是指被研究的社会经济现象所包括的地区范围、隶属范围、分组范围等。时间数列中各个指标所包括的总体范围应前后一致，若某一地区的行政区划发生变化，则该地区的总人口、土地面积、工业总产值、粮食总产量等指标的总体范围与以前有所不同，在编制有关指标的时间数列时，必须先将资料调整成总体范围前后一致，再进行动态分析，这样才能正确说明所研究的问题。

3. 指标的经济内容要统一

统计指标数值反映一定的客观经济内容，因此，时间数列中的各项指标数值所包含的经济现象必须是同质的。例如，要反映某一国有工业企业发展变化的情况时，可编制工业增加值数列，但不能将工业总产值与工业增加值混在一起编制。因为二者经济内容不同，没有可比性，否则会得出错误的结论。随着我国经济体制改革的不断深化，某些指标的经济含义也在变化，因此，保证各期指标经济内容的一致性就十分必要。

4. 各项指标数值的计算方法、计算口径和计量单位要一致

在指标名称及其经济内容一致的前提下，各项指标的计算口径、计量单位和计算方法也应当统一。例如，在编制不同时期第一、二、三产业增加值的时间数列时，应采用统一的不变价格计算，消除价格因素的影响，以便真实地反映各产业的发展情况；在研究企业劳动生产率的变动时，分子产量指标是用实物量还是用价值量，分母人数指标是用全部职工人数还是用生产工人人数，前后均应统一起来。

7.2 时间数列的水平指标

在编制时间数列基础上，为了分析研究社会经济现象在不同时间条件下的发展变化规律，需要计算各种动态分析指标。时间数列的分析指标包括水平指标和速度指标两大类。水平指标包括发展水平指标和平均发展水平、增长量和平均增长量；速度指标包括发展速度和平均发展速度、增长速度和平均增长速度等。水平分析是速度分析的基础，速度分析是水平分析的深入。本节介绍的是时间数列的水平指标。

7.2.1 发展水平

发展水平（time-series data）是时间数列中具体时间条件下的指标数值，具体反映某种社会经济现象在各个时期所达到的规模 and 发展的速度，也叫时间数列水平。它是计算其他动态分析指标的基础，一般用 a_i 表示。

发展水平可以表现为总量指标，也可以表现为相对指标或平均指标。

发展水平按在时间数列中的位置不同来划分，有最初水平、最末水平和中间水平。在一个时间数列中，数列的第一项指标称为最初水平，用符号 a_0 表示；最后一项指标称为最末水平，用符号 a_n 表示；其余就是中间发展水平，用符号 a_1, a_2, \dots, a_{n-1} 来表示。

发展水平按在时间数列分析中的作用不同来划分，有报告期水平与基期水平。我们将被研究时期的发展水平称为报告期水平或计算期水平，通常用 a_i 表示；将作为比较时期的发展水平称为基期水平或基础水平，用 a_0 或 a_{i-1} 表示。

在实际工作中，发展水平常用文字“增加到”、“增加为”、“降低到”、“降低为”表示。

7.2.2 平均发展水平

平均发展水平又称为序时平均数或动态平均数，它是将整个时间数列作为一个整体，从而反映这个整体的一般水平，即将动态数列不同时间上的发展水平加以平均而得到的平均数。

1. 序时平均数与一般平均数的区别与联系

序时平均数与前面所讲的一般平均数有共同之处，都是将现象总体的个体数量差异抽象化，反映社会经济现象的一般水平。但二者却有区别，主要表现在：

（1）序时平均数平均的是事物在不同时间上的数量差异；算术平均数平均的是总体各单位某一数量标志在同一时间上的数量差异。

（2）序时平均数是从动态上说明某一事物在不同时间上发展的一般水平；算术平均数是从静态上说明同一事物总体的不同单位在同一时间上的一般水平。

（3）序时平均数是根据时间数列计算的；算术平均数是根据变量数列计算的。

在动态分析中，利用序时平均数（平均发展水平）来分析社会经济现象的动态变化有很重要的作用：首先，它可以反映社会经济现象在一段时间内发展变化的一般水平，并对其做出概括的说明；其次，利用序时平均数可以消除社会经济现象在短期内所受波动的影响，便于在更广泛的范围内进行对比，可以观察现象的发展趋势；另外，利用序时平均数还可以对某一段时间内某一事物发展达到的一般水平进行不同单位、不同地区之间的比较。

2. 平均发展水平的计算方法

1) 根据总量指标时间数列计算平均发展水平

总量指标时间数列分为时期数列与时点数列两种。计算平均发展水平，需要根据时间数列不同特点采用不同的方法，现分别讨论如下。

① 根据时期数列计算平均发展水平。

由于时期数列中的各项指标数值都是反映社会经济现象在一定时期内的过程总量，具有可加性，因此我们可以采用简单算术平均的方法计算平均发展水平，即将时期数列中研究范围内的各项指标数值之和除以时期项数来得到。计算公式为：

$$\bar{a} = \frac{a_1 + a_2 + \cdots + a_n}{n} = \frac{\sum a}{n} \tag{7-1}$$

式中： \bar{a} ——平均发展水平；

a_i ($i=1,2,3,\cdots,n$)——发展水平；

n ——项数。

【实例 7.1】 某公司 2015 年四个季度销售额资料如表 7.6 所示，现要求计算该公司各季度的平均销售额。

表 7.6 某公司 2015 年销售资料 单位：万元

时 期	第一季度	第二季度	第三季度	第四季度
销售额	450	400	480	500

解：在该例中，由于销售额是时期指标，研究范围是 2015 年全年，所以可采用简单算术平均方法，即该公司 2015 年各季度平均销售额为：

$$\begin{aligned} \bar{a} &= \frac{a_1 + a_2 + \cdots + a_n}{m} = \frac{\sum a}{n} \\ &= \frac{450 + 400 + 480 + 500}{4} = 457.5 \text{（万元）} \end{aligned}$$

② 根据时点数列计算平均发展水平。

由于时点数列中的各项指标数值都是社会经济现象在某一具体时点条件下的瞬间水平，要计算其平均数，就必须知道在每一时点上的指标数值，而事实上这是不可能的，所以在实际中都是有一定的时间间隔。根据资料的掌握情况及排列情况，时点数列又可分为连续时点数列与

间断时点数列。所谓连续时点数列，是在实践中通常用“天”作为最小的时点单位，如果资料逐日登记，我们就将它看成是连续时点数列；如果资料是通过间隔一定时期登记取得的，通常是期初或期末（如月初或月末、季初或季末、年初或年末）登记取得的，我们就将它看成是间断时点数列。

a. 根据连续时点数列计算平均发展水平。

对于连续时点数列分为两种情况：一种情况是资料未经整理，逐日登记，又是逐日排列的，称为间隔相等的连续时点数列。

间隔相等的连续时点数列的平均发展水平的计算公式为：

$$\bar{a} = \frac{\sum a}{n} \quad (7-2)$$

式中： a ——时点指标数值；

n ——天数。

【实例 7.2】 某公司业务一科 9 月上旬每日出勤人数分别为 14、15、14、13、15、14、14、13、15、13。计算 9 月上旬平均每日出勤人数。

$$\text{解：} \bar{a} = \frac{\sum a}{n} = \frac{14+15+14+13+15+14+14+13+15+13}{10} = 14 \text{ (人)}$$

另一种情况是数列中的各项指标并非逐日变动，是在发生变动时才加以登记，称为间隔不等的连续时点数列。此时，资料经过整理，形成频数分布，即存在若干天指标数值相同，则需要以指标数值为变量，以持续天数为权数进行加权算术平均。

间隔不等的连续时点数列的平均发展水平的计算公式为：

$$\bar{a} = \frac{\sum af}{\sum f} \quad (7-3)$$

式中： f 为各项指标的时间间隔，其余符号与前面公式相同。

【实例 7.3】 某企业 2016 年 3 月成品库存资料记录如表 7.7 所示。

表 7.7 某企业 2016 年 3 月成品库存资料

单位：台

时 间	1 日	10 日	15 日	25 日	26 日	28 日
库存量	100	240	400	300	200	200

现要求计算该企业 3 月份成品的平均库存量。

解：从表 7.7 中可以看出，库存量虽然表现为间隔资料（即不是逐日排列），但实质上是具有逐日的详细资料。如 1 日库存量为 100 台，2~9 日虽没有列出，实质上是 2~9 日的库存量没有变化，仍保持 100 台，直到 10 日入库 140 台，变为 240 台。在计算该企业 3 月份平均成品库存量时，应以库存量为变量，以持续天数为权数进行加权算术平均，即：

$$\begin{aligned}\bar{a} &= \frac{\sum af}{\sum f} \\ &= \frac{100 \times 9 + 240 \times 5 + 400 \times 10 + 300 \times 1 + 200 \times 2 + 200 \times 4}{31} \\ &= 245.16 \text{ (台)}\end{aligned}$$

b. 根据间断时点数列计算平均发展水平。

对于间断时点数列分为两种情况：一种情况是时点数列间隔相等，称为间隔相等的间断时点数列。

间隔相等的间断时点数列的平均发展水平的计算公式为：

$$\bar{a} = \frac{\frac{a_1}{2} + a_2 + a_3 + \cdots + a_{n-1} + \frac{a_n}{2}}{n-1} \tag{7-4}$$

式中： $n-1$ ——间断数目。

利用这种方法计算平均发展水平有一个前提条件，即假定现象在相邻两个时点之间的发展变动是均匀的。首先以每一小段的中间值代表该小段的平均水平，然后将各小段的平均水平再用简单算术平均法计算，得到整个被研究时期的平均发展水平，该方法称为“首尾折半法”或“首末折半法”。

【实例 7.4】 某公司 2016 年第一季度员工人数资料如表 7.8 所示。

表 7.8 某公司 2016 年第一季度员工人数统计表

单位：人

时 间	1 月 1 日	2 月 1 日	3 月 1 日	4 月 1 日
员工人数	150	160	156	162

计算第一季度平均员工人数。

解：第一季度包括 1、2、3 三个月，要求该季平均人数，则应先按月求出这三个月的平均人数，然后再对各月平均人数进行平均即为所求。要计算月平均人数，严格地说，必须具有月内每日的人数资料。由前面讨论得知，在实践中，一般是通过期初、期末登记取得资料，而并非逐日登记。所以，只能采用假定的方法推算月平均人数，把月底的人数看成是下月初的人数，并假定月中的变动是均匀的，于是便得到：

$$\text{月平均人数} = \frac{\text{月初人数} + \text{月末人数}}{2}$$

$$1 \text{ 月份人数} = \frac{150 + 160}{2} = 155 \text{ (人)}$$

$$2 \text{ 月份人数} = \frac{160 + 156}{2} = 158 \text{ (人)}$$

$$3 \text{ 月份人数} = \frac{156+162}{2} = 159 \text{ (人)}$$

该公司第一季度平均员工人数为：

$$\begin{aligned}\bar{a} &= \frac{(150+160)/2 + (160+156)/2 + (156+162)/2}{4-1} \\ &= \frac{150/2 + 160 + 156 + 162/2}{3} \\ &= 157.3 \text{ (人)}\end{aligned}$$

另一种情况是时点数列间隔不等，称为间隔不等的间断时点数列。

间隔不等的间断时点数列的平均发展水平的计算公式为：

$$\bar{a} = \frac{\frac{a_1+a_2}{2}f_1 + \frac{a_2+a_3}{2}f_2 + \cdots + \frac{a_{n-1}+a_n}{2}f_{n-1}}{f_1+f_2+\cdots+f_{n-1}} \quad (7-5)$$

式中： f ——各时点的间隔长度。

利用这种方法计算平均发展水平，是以各期期末资料不全为条件，以资料欠缺的时点间隔长度为权数而采用的加权算术平均法。即先求各间隔内的平均数，然后再求总平均数。

注意：上式中指标数值项数为 n ，间隔为 $n-1$ 个。间隔之和 $f_1+f_2+\cdots+f_{n-1}$ 不能写成 $\sum f$ 。

【实例 7.5】 某公司 2015 年度职工人数资料如表 7.9 所示。

表 7.9 某公司 2015 年度职工人数统计表

单位：人

时 间	1 月 1 日	5 月 1 日	8 月 1 日	11 月 31 日	12 月 31 日
职工人数	400	420	426	432	440

计算 2015 年度该公司年平均职工人数。

$$\begin{aligned}\text{解：}\bar{a} &= \frac{\frac{400+420}{2} \times 4 + \frac{420+426}{2} \times 3 + \frac{426+432}{2} \times 4 + \frac{432+440}{2} \times 1}{4+3+4+1} \\ &= \frac{5\,061}{12} = 421.75 \text{ (人)}\end{aligned}$$

由于根据间断时点数列计算平均发展水平是以一定程度假设性为前提的，即假设社会经济现象在各时点间是均匀变化的。因此，在实际中间断时点数列的时间间隔不宜过长，否则准确度较差。

2) 根据相对指标时间数列计算平均发展水平

相对指标时间数列是由具有相互联系的两个总量指标时间数列对比构成的，且作为对比基础的总量指标一般不相等。因此，根据相对指标时间数列来计算平均发展水平时，其基本方法

步骤是：首先计算出分子数列和分母数列的平均发展水平；其次再将这两个平均发展水平进行对比。其基本计算公式为：

$$\bar{c} = \frac{\bar{a}}{\bar{b}} \tag{7-6}$$

式中： \bar{c} ——相对指标时间数列的平均发展水平；

\bar{a} ——分子数列的平均发展水平；

\bar{b} ——分母数列的平均发展水平。

在实际工作中，相对指标时间数列可以由两个时期数列对比形成，也可以由两个时点数列对比形成，还可以由一个时期数列与一个时点数列对比形成。各种数列性质不同，其计算相对指标时间数列的平均发展水平的方法也不一样，具体来看有三种情况。

第一种情况：根据两个时期数列对比所形成的相对指标时间数列来计算平均发展水平。其计算公式为：

$$\bar{c} = \frac{\bar{a}}{\bar{b}} = \frac{\frac{\sum a}{n}}{\frac{\sum b}{n}} = \frac{\sum a}{\sum b} \tag{7-7}$$

【实例 7.6】 某商业企业 2015 年 4、5、6 月份销售额计划完成情况的资料如表 7.10 所示。

表 7.10 某商业企业第二季度销售额计划完成情况				单位：万元
时 间	4 月份	5 月份	6 月份	
实际完成销售额 a	125.6	136.7	197.8	
计划完成销售额 b	115.0	128.0	176.0	
销售额计划完成程度 c (%)	109.2	106.8	112.4	

计算第二季度销售额计划完成程度。

解：销售额计划完成程度的时间数列是相对指标时间数列，是由实际销售额和计划销售额这两个时期数列进行对比计算出来的。由于时期数列的各项指标可以相加，反映更长时间内的现象总量，因此该商业企业第二季度平均销售额计划完成程度为：

$$\bar{c} = \frac{\bar{a}}{\bar{b}} = \frac{\frac{\sum a}{n}}{\frac{\sum b}{n}} = \frac{\sum a}{\sum b} = \frac{125.6+136.7+197.8}{115.0+128.0+176.0} = \frac{460.1}{419.0} = 109.8\%$$

第二种情况：根据两个时点数列进行对比所形成的相对指标时间数列来计算平均发展水平。这种平均数一般是根据间隔相等的间断时点数列所组成的时间数列来计算的。

其平均发展水平的计算公式为：

$$\bar{c} = \frac{\bar{a}}{\bar{b}} = \frac{\frac{a_1 + a_2 + a_3 + \cdots + a_{n-1} + a_n}{2}}{\frac{b_1 + b_2 + b_3 + \cdots + b_{n-1} + b_n}{2}} \tag{7-8}$$

【实例 7.7】 某投资咨询公司的职工人数资料如表 7.11 所示。

月 份	6 月末	7 月末	8 月末	9 月末
投资咨询业务人数	645	670	695	710
全部职工人数	805	826	830	854
投资咨询业务人员占全部职工人数的比重（%）	80.1	83.7	83.7	83.1

试计算第三季度投资咨询业务人员占全部职工人数的比重。

解：上述资料是时间间隔相等的间断时点数列，要计算该公司第三季度投资咨询业务人员占全部职工人数的平均比重，必须先计算出第三季度平均投资咨询业务人员人数和平均全部职工人数，然后再把它们相对比进行求解。

$$\begin{aligned} \bar{c} = \frac{\bar{a}}{\bar{b}} &= \frac{\frac{a_1 + a_2 + a_3 + \cdots + a_{n-1} + a_n}{2}}{\frac{b_1 + b_2 + b_3 + \cdots + b_{n-1} + b_n}{2}} \\ &= \frac{\frac{645}{2} + 670 + 695 + \frac{710}{2}}{\frac{805}{2} + 826 + 830 + \frac{854}{2}} = \frac{2\,042.5}{2\,485.5} = 82.18\% \end{aligned}$$

第三种情况：根据一个时期数列与另一个时点数列进行对比所形成的相对指标时间数列来计算平均发展水平，其计算公式为：

$$\bar{c} = \frac{\bar{a}}{\bar{b}} = \frac{(a_1 + a_2 + a_3 + \cdots + a_{n-1} + a_n) / n}{\left(\frac{b_0}{2} + b_1 + b_2 + b_3 + \cdots + b_{n-1} + \frac{b_n}{2}\right) / n} \tag{7-9}$$

【实例 7.8】 某超市第四季度商品销售额与月末库存额资料见表 7.12 所示。

月 份	9 月	10 月	11 月	12 月
商品销售额 <i>a</i>	75	80	150	240
月末库存额 <i>b</i>	35	45	55	65
商品流转次数 <i>c</i> （次）	—	2	3	4

试计算第四季度平均每月商品流转次数。

解：商品流转次数时间数列是一个相对指标时间数列，构成相对数列的分子是商品销售额时期数列，分母是月末库存额时点数列，所以应该分别按时期数列计算分子平均发展水平，按

间隔相等的间断时点数列计算分母平均发展水平,然后再将分子与分母对比,求得相对指标时间数列的平均发展水平。

第四季度平均每月商品流转次数为:

$$\begin{aligned}\bar{c} &= \frac{\bar{a}}{\bar{b}} = \frac{(a_1 + a_2 + a_3 + \cdots + a_{n-1} + a_n) / n}{\left(\frac{b_0}{2} + b_1 + b_2 + b_3 + \cdots + b_{n-1} + \frac{b_n}{2} \right) / n} \\ &= \frac{\frac{80 + 150 + 240}{3}}{\frac{\frac{35}{2} + 45 + 55 + \frac{65}{2}}{3}} \\ &= \frac{156.7}{50} \\ &= 3.13 \text{ (次)}\end{aligned}$$

3) 根据平均指标时间数列计算平均发展水平

平均指标时间数列由一般平均数和序时平均数所组成。由于这两种平均数各有不同的特点,因此,由它们组成的时间数列计算平均发展水平的方法也不相同。

第一种情况:根据一般平均数组成的时间数列来计算平均发展水平。一般平均数时间数列,实质上也是由两个总量指标时间数列对比所形成的,因此,其平均发展水平的计算与相对指标时间数列计算平均发展水平的方法相同。

第二种情况:根据序时平均数组成的时间数列来计算平均发展水平。根据间隔相等的序时平均数所组成的平均指标时间数列计算平均发展水平,可直接采用简单算术平均法计算,公式见(7-10)。而由间隔不等的序时平均数所组成的平均指标时间数列计算平均发展水平,则以间隔长度为权数进行加权计算,公式见(7-11)。

$$\bar{c} = \frac{\sum \bar{a}}{n} \quad (7-10)$$

$$\bar{c} = \frac{\sum \bar{a}f}{\sum f} \quad (7-11)$$

7.2.3 增长量

增长量又称增减量,是时间数列中报告期水平与基期水平之差,它说明社会经济现象在一定时期内增减变化的绝对量。其计算公式为:

$$\text{增长量} = \text{报告期水平} - \text{基期水平}$$

若计算结果为正值,表示增加量;若计算结果为负值,表示减少量。由于对比选择的基期

不同，增长量可分为逐期增长量和累计增长量两种。

逐期增长量是报告期水平与上一期水平之差，表明各报告期水平较上一时期水平逐期增减变动的绝对数量，即逐期增长量=报告期水平-上一期水平；累计增长量是报告期水平与某一固定基期水平之差，表明各报告期水平较某一固定时期水平增减变动的绝对数量，即累计增长量=报告期水平-某一固定基期水平。以符号表示为：

逐期增长量： $a_1-a_0,a_2-a_1,a_3-a_2,\cdots,a_n-a_{n-1}$

累计增长量： $a_1-a_0,a_2-a_0,a_3-a_0,\cdots,a_n-a_0$

逐期增长量与累计增长量之间有一定的数量关系，即累计增长量等于相应各个逐期增长量之和，用公式表示为：

$$a_n-a_0=(a_1-a_0)+(a_2-a_1)+(a_3-a_2)+\cdots+(a_n-a_{n-1}) \tag{7-12}$$

【实例 7.9】 我国 2009—2014 年货物进出口总额的逐期增长量与累计增长量如表 7.13 所示。

表 7.13 我国 2009—2014 年货物进出口总额

年 份		2009	2010	2011	2012	2013	2014
货物进出口总额（亿元）		150 648	201 722	236 402	244 160	258 169	264 334
增长量（亿元）	累计	0	51 074	85 754	93 512	107 521	113 686
	逐期	—	51 074	34 680	7 758	14 009	6 165
发展速度（%）	定基	100	133.90	156.92	162.07	171.37	175.46
	环比	—	133.90	117.19	103.28	105.74	102.39
增长速度（%）	定基	0	33.90	56.92	62.07	71.37	75.46
	环比	—	33.90	17.19	3.28	5.74	2.39
增长 1%的绝对值（亿元）		—	1 506.48	2 017.22	2 364.02	2 441.60	2 581.69

此外，对于受季节因素影响较明显的社会经济指标，为了排除季节变化的影响，还可计算年距增长量。它是报告期为某月（季）水平与上年同月（季）水平之差，说明经济现象经过一年增减变化的绝对量。其计算公式为：

$$\text{年距增长量}=\text{本期发展水平}-\text{上年同期发展水平} \tag{7-13}$$

【实例 7.10】 某商场 2015 年第一季度商品销售额为 8 800 万元，2016 年第一季度商品销售额为 9 200 万元，求年距增长量。

解：2015 年第一季度年距增长量=9 200-8 800=400（万元）

7.2.4 平均增长量

平均增长量，是指逐期增长量的简单算术平均数，它说明经济现象在一段较长时间内，每期平均增减变化的数量。其计算公式为：

$$\text{平均增长量} = \frac{\text{逐期增长量之和}}{\text{逐期增长量项数}} = \frac{\text{累计增长量}}{\text{数列项数} - 1} \quad (7-14)$$

【实例 7.11】 根据表 7.13 的资料, 计算我国货物进出口总额在 2009—2014 年间平均每年增长量。

解: 平均增长量 = $(51\,074 + 34\,680 + 7\,758 + 14\,009 + 6\,165) / 5 = 22\,737.2$ (亿元)

或平均增长量 = $113\,686 / 5 = 22\,737.2$ (亿元)

7.3 时间数列的速度指标

表示社会经济现象发展速度的指标有: 发展速度、增长速度、平均发展速度和平均增长速度。

7.3.1 发展速度

发展速度 (rate of expansion) 是反映客观现象发展变化快慢程度的相对指标, 它是根据报告期水平与基期水平之比求得的, 表明报告期水平已发展到基期水平的若干倍或百分之几, 其计算公式为:

$$\text{发展速度} = \frac{\text{报告期水平}}{\text{基期水平}} \quad (7-15)$$

由于对比所采用的基期不同, 发展速度指标可分为定基发展速度和环比发展速度两种。

定基发展速度是报告期水平与某一固定时期的水平之比, 表明报告期水平对某一固定时期水平已发展到若干倍或百分之几, 表明客观现象在较长时期内总的发展速度, 其计算公式为:

定基发展速度 = 报告期水平 / 固定基期水平

符号表示为: $\frac{a_1}{a_0}, \frac{a_2}{a_0}, \frac{a_3}{a_0}, \dots, \frac{a_{n-1}}{a_0}, \frac{a_n}{a_0}$

例如, 表 7.13 中货物进出口总额的定基发展速度就是 2009—2014 年各期的发展水平分别与 2009 年的发展水平对比而得出的。

环比发展速度是报告期水平与前期水平之比, 说明报告期水平对前期水平已发展到若干倍或百分之几, 表明客观现象在两个相邻时期或时点上的发展速度, 其计算公式为:

环比发展速度 = 报告期水平 / 前期水平

符号表示为: $\frac{a_1}{a_0}, \frac{a_2}{a_1}, \frac{a_3}{a_2}, \frac{a_4}{a_3}, \dots, \frac{a_n}{a_{n-1}}$

例如, 表 7.13 中货物进出口总额的环比发展速度就是 2009—2014 年各期的发展水平分别与前一年的发展水平对比而得出的。

定基发展速度与环比发展速度虽然在说明客观现象时的侧重点有所不同，但它们之间存在以下两种换算关系。

(1) 定基发展速度等于相应时期内各个环比发展速度的连乘积，即

$$\frac{a_n}{a_0} = \frac{a_1}{a_0} \times \frac{a_2}{a_1} \times \frac{a_3}{a_2} \times \frac{a_4}{a_3} \times \cdots \times \frac{a_n}{a_{n-1}} \quad (7-16)$$

【实例 7.12】 由表 7.13 得出：

2014 年的定基发展速度 = $(1.3390 \times 1.1719 \times 1.0328 \times 1.0574 \times 1.0239) = 175.46\%$

(2) 相邻的两个定基发展速度相除，得出相应的环比发展速度，即

$$\frac{a_i}{a_0} \div \frac{a_{i-1}}{a_0} = \frac{a_i}{a_{i-1}} \quad (7-17)$$

【实例 7.13】 由表 7.13 得出：

2014 年的环比发展速度 = $175.46\% / 171.37\% = 102.39\%$ 。

在实际工作中，可以根据以上两种换算关系，对发展速度进行相互推算。

为了消除季节因素对社会经济现象发展变化的影响，在计算月份或季度发展速度时，可选用上年同期作为对比的基期，计算年距发展速度。此外，还可以选用历史最高水平的时间作为对比的基期，以反映在报告期已经发展到或不及历史最高水平的程度。

7.3.2 增长速度

增长速度 (rate of growth) 又称增减速度，是反映客观现象增长快慢程度的相对指标，它是增长量与基期发展水平之比，用以说明报告期水平比基期水平增长了若干倍或百分之几，其计算公式为：

$$\begin{aligned} \text{增长速度} &= \frac{\text{增长量}}{\text{基期水平}} = \frac{\text{报告期水平} - \text{基期水平}}{\text{基期水平}} \\ &= \text{发展速度} - 1 \end{aligned} \quad (7-18)$$

从上式可以看出，当发展速度大于 1 或 100% 时，增长速度为正值，表示现象的增长程度；当发展速度小于 1 或 100% 时，增长速度为负值，表示现象的降低程度。

增长速度由于所采用的基期不同，也可分为定基增长速度和环比增长速度。

定基增长速度是累积增长量与固定时期发展水平之比，它表明客观现象从某一固定基期至报告期这一时期内的增长速度。其计算公式为：

$$\begin{aligned} \text{定基增长速度} &= \frac{a_i - a_0}{a_0} = \frac{a_i}{a_0} - 1 \\ &= \text{定基发展速度} - 1 \quad (i=1, 2, 3, \cdots, n) \end{aligned} \quad (7-19)$$

例如,表 7.13 中货物进出口总额的定基增长速度就是 2009—2014 年累计增长量除以固定基期(2009 年)水平计算,也可以由定基发展速度减去 1 而求得。

环比增长速度是逐期增长量与前期发展水平之比,它表明客观现象逐期增长的速度,其计算公式为:

$$\begin{aligned}\text{环比增长速度} &= \frac{a_i - a_{i-1}}{a_{i-1}} = \frac{a_i}{a_{i-1}} - 1 \\ &= \text{环比发展速度} - 1 \quad (i=1,2,3,\cdots,n) \quad (7-20)\end{aligned}$$

例如,表 7.13 中货物进出口总额的环比增长速度就是 2009—2014 年各年的逐期增长量除以前一期水平计算,也可以由环比发展速度减去 1 而求得。

值得注意的是,定基增长速度和环比增长速度之间没有量的乘除关系,所以不能直接进行换算。增长速度可以计算年距增长速度,也可以计算报告期水平与历史最高水平的差异程度。

为了反映增长速度的实际效果,有时需要计算每增长百分之一的绝对值指标,它表明速度每增长 1%,发展水平指标在绝对数上增长了多少。其计算公式为:

$$\begin{aligned}\text{增长1\%的绝对值} &= \frac{\text{逐期增长量}}{\text{环比增长速度} \times 100} \\ &= \frac{\text{前一期水平}}{100} \quad (7-21)\end{aligned}$$

例如,表 7.13 中货物进出口总额每年增长 1%绝对值是不同的,2010 年为 1 506.48 亿元,2014 年则为 2 581.69 亿元。可见,运用增长 1%绝对值指标,实际上是遵循了相对指标要与总量指标相结合运用的原则,把增长速度和增长量结合起来观察、分析问题是必要的。

7.3.3 平均发展速度

平均发展速度是各个时期环比发展速度的序时平均数,说明客观现象在一个较长的时间内逐年平均发展变化的平均程度。平均发展速度也是一种动态平均数,但不能照搬前面计算序时平均数的方法来计算。在实际统计工作中,平均发展速度的计算方法有水平法(几何平均法)和累积法(方程式法)两种。

1. 水平法

水平法又称几何平均法。平均发展速度是总速度的平均,且各环比发展速度的连乘积等于总速度,因此,计算现象平均发展速度时不能用算术平均法,而要用几何平均法计算。其计算公式为:

$$\bar{x} = \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n} = \sqrt[n]{\prod x} \quad (7-22)$$

式中: \bar{x} ——平均发展速度;

x ——各期环比发展速度;

n ——环比发展速度的项数;

\prod ——连乘符号。

由于动态数列中定基发展速度等于各环比发展速度的连乘积,所以,计算平均发展速度的公式又可以表示为:

$$\bar{x} = \sqrt[n]{\frac{a_1}{a_0} \times \frac{a_2}{a_1} \times \cdots \times \frac{a_n}{a_{n-1}}} = \sqrt[n]{\frac{a_n}{a_0}} \quad (7-23)$$

一定时期的定基发展速度即为总速度。如果用 R 表示总速度,则平均发展速度的公式还可以表示为:

$$\bar{x} = \sqrt[n]{R} \quad (7-24)$$

以上计算平均发展速度的三个公式,虽然形式不同,但其实质与计算结果完全相同。计算平均发展速度究竟采用哪个公式,主要取决于所掌握的资料。如果掌握的资料是各年的环比发展速度,用式(7-22);如果掌握的资料是最初水平和最末水平或各期发展水平,用式(7-23);如果已知本期的定基发展总速度,则用式(7-24)。

【实例 7.14】 根据表 7.13 的资料,已知我国货物进出口总额 2009—2014 年各年的环比发展速度分别为 133.90%、117.19%、103.28%、105.74%、102.39%,求年平均发展速度。

$$\text{解: } \bar{x} = \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n} = \sqrt[n]{\prod x} = 111.90\%$$

根据表 7.13,已知我国货物进出口总额 2009 年为 150 648 亿元,2014 年为 264 334 亿元,即 $a_n=264\,334$, $a_0=150\,648$, $n=5$,则其年平均发展速度为:

$$\bar{x} = \sqrt[n]{\frac{a_n}{a_0}} = 111.90\%$$

根据表 7.13,已知我国货物进出口总额 2009—2014 年的总发展速度为 175.46%,其年平均发展速度为:

$$\bar{x} = \sqrt[n]{R} = 111.90\%$$

计算结果表明,应用以上三种公式对同一现象来计算平均发展速度,所得出的结果是一致的(有时出现小数不一致的现象,属于计算过程中四舍五入造成的)。

【实例 7.15】 根据表 7.13 的资料,如果以 2014 年我国货物进出口总额为基数,其后每年以 111.90% 的速度发展,到 2017 年我国货物进出口总额将达到多少亿元?

$$\text{解: } a_n = a_0 \times \bar{x}^n = 264\,334 \times 1.119^3 = 370\,376 \text{ (亿元)}$$

即按 111.90% 的速度发展,到 2017 年我国货物进出口总额将达到 370 376 亿元。

【实例 7.16】 根据表 7.13 的资料, 若每年以 111.90% 的速度发展, 到哪一年货物进出口总额将达到 400 000 亿元?

解: 根据公式 (7-24), 可推导出 $n = \frac{\ln R}{\ln \bar{x}} = 3.684$ (年)

即 2014 年后 3.684 年, 亦即到 2017 年 9 月左右, 货物进出口总额将达到 400 000 亿元。

用水平法计算平均发展速度具有两个特点:

第一, 这种方法侧重于考察最末一期的发展水平。

第二, 这种方法不能准确反映中间水平的起伏状况。从理论上讲, 用水平法计算的平均发展速度, 是对一定发展阶段各期环比发展速度的平均, 受各个时期发展水平的影响; 但从计算公式观察, 它只突出了最初水平和最末水平的影响, 却不能全面反映在整个发展阶段各期发展快慢的差别。

因此, 在运用这一指标时, 应注意最初水平与最末水平是否受特殊因素影响。同时, 要联系各期环比发展速度加以分析, 既要水平, 也要看速度, 必要时用分段平均发展速度补充总平均发展速度, 以便对现象的发展做出全面客观的评价。

2. 累积法

累积法又称方程法。这种方法的特点是: 从最初水平 a_0 开始, 每期以平均发展速度 \bar{x} 发展, 从理论上计算的各期水平之和应等于实际各期水平 a_i ($i=1, 2, \dots, n$) 之和, 即

$$\bar{x} + \bar{x}^2 + \dots + \bar{x}^n = \frac{\sum_{i=1}^n a_i}{a_0} \quad (7-25)$$

该高次方程的正根就是所要求的平均发展速度 \bar{x} 。一般借助计算机可以求解, 在实际工作中也可查《平均增长速度查对表》来求得。具体步骤如下:

① 计算各期定基发展速度之和, 即求 $\sum_{i=1}^n a_i / a_0$ 。

② 用 $\sum_{i=1}^n a_i / a_0$ 再除以时期数 n , 判断平均增长速度的正负性, 如果除得的结果大于 1, 表示各期的定基发展速度平均在 100% 以上, 则该数列属于递增型; 反之, 除得的结果小于 1, 表示各期的定基发展速度平均在 100% 以下, 则该数列则属于递减型。

③ 根据各期定基发展速度之和查表, 即可得平均增长速度和平均发展速度。

【实例 7.17】 某乡镇企业 2009 年固定资产投资额为 122.3 万元, 2010—2014 年间累积投资达 701.5 万元, 计算该企业 2009—2014 年间固定资产投资的平均发展速度。

解: 已知 $a_0 = 122.3$ 万元, $\sum_{i=1}^n a_i = 701.5$ 万元

则得 $\sum_{i=1}^n a_i / a_0 = 702.5 / 122.3 = 5.7359$ ，即 573.59%

且 $5.7359 / 5 = 1.14718 > 1$

所以该数列为递增型。

查表 7.14，在 $n=5$ 一栏内，找接近 573.59% 的数字为 573.38%，该数所对应的平均每年增长速度为 4.6%，即为该企业 2009—2014 年间的固定资产投资平均增长速度，平均发展速度则为 104.6%。

表 7.14 累积法查对表

平均每年增长%	各年发展水平总和为基期的%				
	1 年	2 年	3 年	4 年	5 年
∴	∴	∴	∴	∴	∴
4.4	104.40	213.39	327.18	445.98	570.01
4.5	104.50	213.70	327.82	447.07	571.69
4.6	104.60	214.01	328.46	448.17	573.38
4.7	104.70	214.32	329.09	449.26	575.08
∴	∴	∴	∴	∴	∴

用于计算平均发展速度的水平法和累积法各自具有不同的特点和侧重面，计算平均发展速度究竟采用哪种方法，应根据所研究现象的具体情况来确定。一般情况下，如果所研究现象侧重于考察最末一期所达到的水平，则采用水平法计算平均发展速度，如产量、工资等；如果所研究现象侧重于考察一段较长时间内的总量，则宜采用累积法计算平均发展速度，如基建投资、占用土地等。

7.3.4 平均增长速度

平均增长速度说明客观现象在一个较长的时间内逐年平均增长变化的程度。平均增长速度和平均发展速度的关系是：

平均增长速度=平均发展速度-1 (7-26)

但在实际工作中，平均增长速度不能根据各期的增长速度或总增长速度来计算，而是通过先计算平均发展速度，再依据两者之间关系进行换算。

【实例 7.18】 某生产公司产品单位成本在前五年环比增减速度资料如表 7.15 所示，求平均增长速度。

表 7.15 某生产公司产品单位成本 2011—2015 年环比增减速度

年 份	2010	2011	2012	2013	2014	2015
环比增长速度%	—	-1.0	-1.5	-2.0	-1.8	-0.5

解：首先利用发展速度与增长速度之间的关系，求出各年环比发展速度，然后据此计算平均发展速度，最后得出平均增长速度。

$$\bar{x} = \sqrt[5]{(-1.0\%+1) \times (-1.5\%+1) \times (-2.0\%+1) \times (-1.8\%+1) \times (-0.5\%+1)} = 98.64\%$$

平均增长速度=98.64%-1=-1.36%。

7.4 时间数列趋势分析预测

社会经济现象的发展变化是由许多错综复杂的因素共同作用的结果，在这些因素中，有的是系统因素，它对现象的发展起着决定性作用；有的是偶然因素，它对现象的发展只起到局部的、临时的、非决定性的作用。由于各种因素的作用大小和方向不同，使时间数列呈现出不同的变动形态。时间数列分析的任务就是要正确地确定时间数列的性质，对各构成因素进行分解和测定，并对未来的状态做出推断和预测。作为基本分析，一般来说，把时间数列的构成因素按性质和作用分为四类，即长期趋势、季节变动、循环变动和不规则变动。

长期趋势（trend）是指序列变动的总的方向性趋势。形成一个序列的长期趋势是带根本性因素的变动趋势，如人口增长、科学技术的不断进步，必然导致社会生产总量呈现增长变动的趋势。这种长期趋势通常可以认为是由各种固定的因素作用于同一方向而形成的，这些固定的因素既有随时间推移按直线变化，也有呈曲线变动。所以，长期趋势是对未来状况进行预测和推断的主要依据。

季节变动（seasonal effects）是指在一年以内，时间数列受自然季节和社会习俗等因素影响而发生的有规律的、周期性的变动，如服装的季节性供应，铁路、公路、航空等客运量在春节前后出现高峰等。引起季节变动的原因既有自然因素，也有人为因素，如气候条件、节假日及风俗习惯等。季节变动的影响有以一年为周期的，也有以一日、一周、一月为周期的。认识和掌握季节变动，对于近期行动的决策具有重要作用。

循环变动（cyclical effects）是指围绕着长期趋势出现的，具有一定循环起伏形态的变动。这种循环起伏与季节变动不同，循环的幅度和周期很不规则，周期长的可达几十年，周期短的一般也在3~5年，如社会经济的繁荣、危机、萧条、复苏的周期变动便是典型的中循环波动，其周期一般在8~9年。循环变动也不同于长期趋势，它所表现的不是单一方向的持续运动，而是涨落相间的波浪式发展。测定循环变动，掌握其发展变化规律，这对于人们认识事物，利用或克服其产生的影响具有重要的意义。如我国统计工作中开展的宏观经济预测、预警系统的研究，有助于及时发现经济波动的趋势。

不规则变动（irregular fluctuations）是指客观现象由于突发事件或偶然因素引起的无规律性的变动，又称剩余变动或随机变动，如政治动荡、大的自然灾害、战争等。它们是时间数列中无法由上述三个因素解释的部分。不规则变动与时间无关。

把时间数列的这些构成因素与时间数列的关系用一定的数学关系式来表示，就是时间数列的分解模式。分解时间数列的基本模式有加法模式和乘法模式。

设时间数列为 y ，长期趋势为 T ，季节变动为 S ，循环变动为 C ，不规则变动为 I ，则：

时间数列分解的加法模式为：

$$y=T+S+C+I \quad (7-27)$$

时间数列分解的乘法模式为：

$$y=T \times S \times C \times I \quad (7-28)$$

值得注意的是，时间数列分析并不能作为对前景预测的唯一依据。因为时间数列中的不规则变动本身就是难以预计的，特别是一些偶然性重大事故的发生，规律更难以预料，即便是较有规律的因素（ T 、 S 、 C ），它们的变动规律也不一定就是过去历史的简单重复。因此，在利用时间数列分析对社会经济现象进行预测时，预测的时间不宜过长，并应注意对一些影响其发展的主要因素进行分析。

在实际工作中，需要根据研究对象的性质、研究的目的和所掌握的资料等具体情况确定采用哪一种模式进行分析。一般来说，在社会经济统计中，主要采用乘法模式，把受各个因素影响的变动分别测定出来，为决策提供依据。事实上，有些现象的动态数列并非4种变动俱在，从长期来看，揭示经济现象发展的长期趋势和测定其受季节变动的影响，对于每一个具体的动态数列来讲都是十分重要的问题。本节阐述长期趋势的测定与预测方法，下一节将阐述季节变动的测定与预测方法。

7.4.1 长期趋势分析预测

长期趋势是研究某种现象在一个相当长的时期内持续向上或向下发展变动的趋势。例如，我国的国内生产总值在社会主义市场经济运作中，呈现不断上升的长期趋势。测定长期趋势的主要目的有：首先，在于把握现象的趋势变化；其次，从数量方面来研究现象发展的规律性，探求合适趋势线，为进行统计预测提供必要条件；最后，测定长期趋势可以消除原有动态数列中长期趋势的影响，以便更好地显示和测定季节变动。

在实际工作中，常常把趋势分析与统计预测结合在一起。趋势分析与统计预测是现代化的管理方法，它可以反映社会经济现象发展变化的规律，从而使我们对未来有比较科学的认识。通过预测为领导机关和管理部门制定正确的决策提供依据。

反映现象发展的长期趋势有两种基本形式：一种是直线趋势；另一种是非直线趋势，即曲线趋势。当所研究现象在一个相当长的时期内呈现出比较一致的上升或下降的变动，如遵循一直线发展，则为直线趋势，可求出一条直线代表之，这条直线也可叫作趋势直线。趋势直线上升或下降，表示这种现象的数值逐年俱增或俱减，且每年所增加或减少的数量大致相同，所以直线趋势的变化率或趋势线的斜率基本上是不变的。而非直线趋势，其变化率或趋势线的斜率是变动的。

研究现象发展的长期趋势，就必须对原来的时间数列进行统计处理，一般称为时间数列修匀，即进行长期趋势测定。测定长期趋势常用的主要方法有时距扩大法、移动平均法、最小平方法。

1. 时距扩大法

这是测定直线趋势的一种简单的方法。当原始时间数列中各指标数值上下波动，使现象变化规律表现不明显时，可通过扩大数列的时间间隔，对原资料加以整理，以反映现象发展的趋势。用时距扩大法修匀时间数列，既可用总量指标表示，也可用平均指标表示。前者仅适用于时期数列；后者既适用于时期数列，也适用于时点数列。

【实例 7.19】 利用时距扩大法修匀指标。某公司 2015 年产量资料如表 7.16 所示。

表 7.16 某公司 2015 年产量（1）

月 份	1	2	3	4	5	6	7	8	9	10	11	12
产量（万件）	50	45.8	52	51.3	50.9	55.7	54.6	58.5	57	59.9	61.2	62.1

从表 7.16 中可看出，该公司各月产量有上升的发展趋势，但月与月之间存在着升降交替的现象。如果将各月的产量合并为按季计算的产量，即扩大时距，则新编的动态数列如表 7.17 所示。

表 7.17 某公司 2015 年产量（2）

季 度	第一季度	第二季度	第三季度	第四季度
产量（万件）	147.8	157.9	170.1	183.2

在时距扩大修匀后的新数列中，产量的变动情况就呈现出明显的上升趋势。

应用时距扩大法来修匀动态数列，应遵循下列基本要求。

（1）同一数列前后时间间隔应当一致，以便于比较。

（2）时间间隔的长短，应根据具体现象的性质和特点而定。如果时距扩大不够，就不能消除现象变动中的偶然因素；反之，如果时距过长，则修匀后整理出来的动态数列指标就少，会掩盖现象发展的具体趋势。

2. 移动平均法

移动平均法（moving average）是对原有时间数列进行修匀，来测定其长期趋势的一种较为简单的方法。这个方法就是采用逐项递推移动的方法，分别计算一系列移动的序时平均数，形成一个新的派生的序时平均数动态数列，来代替原有的时间数列。在这个新的时间数列中，短期的偶然因素引起的变动被削弱了，从而呈现出明显的长期趋势。

【实例 7.20】 表 7.18 为某企业 2006—2015 年的商品销售额资料。

表 7.18 某企业 2006—2015 年的商品销售额

单位：亿元

年 度	商品销售额	三项移动平均	五项移动平均	四项移动平均	四项移动平均正位
2006	9.60	—	—	—	—
2007	10.66	11.26	—	12.14	—
2008	13.52	12.98	12.32	13.00	12.58

续表

年 度	商品销售额	三项移动平均	五项移动平均	四项移动平均	四项移动平均正位
2009	14.76	13.78	13.20	13.84	13.42
2010	13.08	13.94	14.08	14.22	14.04
2011	14.00	14.04	15.04	15.10	14.66
2012	15.04	15.78	15.68	16.32	15.72
2013	18.28	17.1	16.8	17.5	16.92
2014	17.96	18.32	—	—	—
2015	18.70	—	—	—	—

(1) 三项移动平均。

第一个平均数 = $\frac{9.60+10.66+13.52}{3} = 11.26$ ，对正第二项的原值；

第二个平均数 = $\frac{10.66+13.52+14.76}{3} = 12.98$ ，对正第三项的原值。

依次类推，边移动边平均，求得三项移动平均新数列共 8 项。

(2) 五项移动平均。

第一个平均数 = $\frac{9.60+10.66+13.52+14.76+13.08}{5} = 12.32$ ，对正第三项的原值；

第二个平均数 = $\frac{10.66+13.52+14.76+13.08+14.00}{5} = 13.20$ ，对正第四项的原值。

依次类推，边移动边平均，求得五项移动平均新数列共 6 项。

(3) 四项移动平均。

第一个平均数 = $\frac{9.60+10.66+13.52+14.76}{5} = 12.14$ ，对正第二项和第三项的中间；

第二个平均数 = $\frac{10.66+13.52+14.76+13.08}{5} = 13.00$ ，对正第三项和第四项的中间。

依次类推，边移动边平均，求得四项移动平均新数列共 7 项。

由于每个指标数列都和原动态数列错半期，无法直接进行对比，还必须进行一次正位平均。即再进行一次两项移动平均，这样新序时平均数数列的各期才能和原动态数列对准，形成新的四项正位平均指标数列共 6 项。

从表 7.18 的计算结果可以看出，移动平均可以使时间数列中短期的偶然因素弱化，整个数列被修匀得更加平滑，波动趋于平稳。

应用移动平均法分析长期趋势时，应注意下列问题。

第一,用移动平均法对原动态数列修匀,修匀程度的大小与原数列移动平均的项数多少有关。用5项移动平均比3项移动平均修匀程度更大些。也就是说,修匀的项数越多,效果越好,即趋势线越平滑。

第二,移动平均法所取的项数的多少,应视资料的特点而定。原有时间数列如有循环周期,则移动平均的项数以循环周期的长度为准。事实证明,当移动平均的时期长度等于周期长度或其整倍数时,它就能把周期的波动完全抹掉。当数列资料为季资料时,可采用4项移动平均;若根据各年的月份资料,则应取12项移动平均,这样可消除所受季节变动的影响,能较为准确地揭示现象发展的长期趋势。

第三,移动平均法采用奇数项移动比较简单,一次即得趋势值,如表7.18所示。采用偶数项移动平均时,由于偶数项移动平均数都是在两项中间位置,所以要将第一次移动的平均值再进行两项“移正平均”,得出移正值时间数列,以显示出现象变动趋势。由于偶数项移动平均比较复杂,因此,一般常以奇数项为长度。

第四,移动平均后的数列要比原数列项数减少。移动时采用的项数愈多,虽能更好地进行修匀,但所得趋势值的项数就愈少。一般情况下,移动平均项数与趋势值的项数关系为:趋势值项数=原数列项数-移动平均项数+1。因此,为了便于看出现象的发展趋势,要视具体情况,以确定移动平均的项数不宜太多。

3. 最小平方法

最小平方法(least square method)研究现象的发展趋势,就是运用一定的数学模型,对原有的时间数列配合一条适当的趋势线来进行修匀。根据最小平方法的原理,这条趋势线必须满足最基本的要求,即原有数列的实际数值与趋势线的估计数值的离差平方之和为最小。用公式表示如下:

$$\sum (y - y_c)^2 \rightarrow \text{最小值}$$

式中: y_c ——趋势线的估计数值;

y ——原有数列的实际数值。

长期趋势的类型很多,有直线型和曲线型,而最小平方法既可用于配合直线,也可用于配合曲线,所以它是分析长期趋势的十分普遍和理想的方法。下面主要介绍根据社会经济现象的基本趋势,如何用最小平方法配合直线方程、抛物线方程及指数曲线方程。

(1) 直线方程。如果现象的发展,其逐期增长量大体上相等,则可考虑配合直线趋势。直线方程的一般形式为:

$$y_c = a + bt \quad (7-29)$$

式中: y_c ——趋势线的估计数值;

t ——代表动态数列的时间单位;

a 和 b 为参数,分别代表直线的截距和斜率。

根据最小平方法的基本要求，即 $\sum (y-y_c)^2 \rightarrow$ 最小值和多元函数求极值的定理，可用求偏导数的方法，导出以下联立方程组：

$$\begin{aligned}\sum y &= na + b \sum t \\ \sum ty &= a \sum t + b \sum t^2\end{aligned}\quad (7-30)$$

式中： n ——动态数列的项数。

其他符号与前面公式相同。

解联立方程组可得 a 和 b 两个参数值：

$$\begin{aligned}b &= \frac{n \sum ty - \sum t \sum y}{n \sum t^2 - (\sum t)^2} \\ a &= \frac{\sum y - b \sum t}{n} = \bar{y} - b \bar{t}\end{aligned}\quad (7-31)$$

将求得的两个参数代入直线趋势模型中，便可得到与实际观察值相对应的趋势值。由此可以形成一条趋势直线，既可以认识现象的发展变化动态，还可以预测未来。

【实例 7.21】 某公司 2010—2015 年在某地区销售甲产品资料如表 7.19 所示。求出直线趋势方程，并预测 2016 年销售量。

表 7.19 某公司 2010—2015 年在某地区甲产品销售量

单位：千台

年 份	销售量 y	逐期增长量	年份序号 t	t^2	ty	y_c
2010	68	0	0	0	0	63.7
2011	71	3	1	1	71	71.38
2012	75	4	2	4	150	75.46
2013	79	4	3	9	237	79.54
2014	84	5	4	16	336	83.62
2015	88	4	5	25	440	87.7
合 计	465	—	15	55	1 234	465

解：根据表 7.19，可知

$$n=6, \sum t=15, \sum y=456, \sum t^2=55, \sum ty=1\,234$$

将这些数据资料代入上面联立方程，得

$$\begin{aligned}465 &= 6a + 15b \\ 1\,234 &= 15a + 55b\end{aligned}$$

解方程组得： $a=67.3, b=4.08$

配合出的直线趋势方程式为： $y_c=67.3+4.08t$

利用已拟合的直线趋势方程式，预测该地区 2016 年甲产品可能达到的销售量为：

$$y_{2016}=67.3+4.08\times6=91.78 \text{（千台）}$$

为了计算方便，可巧妙地对时间 t 进行取值。当时间项数为奇数时，可假设 t 的中间项为 0，这时时间项依次排列为： $\cdots,-3,-2,-1,0,1,2,3,\cdots$ ；当时间项数为偶数时，时间项依次排列为： $\cdots,-5,-3,-1,1,3,5,\cdots$ 。这样，原点 O 实际上是在数列正中相邻两个时间的中点，保证了时间项间隔相等。以上两种设 t 的方法目的是使 $\sum t=0$ ，此时，上述联立方程组可简化为：

$$\begin{cases} \sum y = na \\ \sum ty = b \sum t^2 \end{cases} \tag{7-32}$$

即：

$$\begin{cases} a = \frac{\sum y}{n} \\ b = \frac{\sum ty}{\sum t^2} \end{cases} \tag{7-33}$$

这样可以大大简化计算程序，因此，通常把这种计算称为简捷法运算。在用最小平方拟合直线方程时，不论时间数列的时间项数是奇数项还是偶数项，都可以运用简捷运算方法来计算。

【实例 7.22】 仍以表 7.19 中的资料为例，用简捷法进行计算，如表 7.20 所示。

表 7.20 最小平方方法简捷法计算表

年 份	销售量 y	逐期增长量	年份序号 t	t^2	ty	y_c
2010	68	0	-5	25	-340	63.7
2011	71	3	-3	9	-213	71.38
2012	75	4	-1	1	-75	75.46
2013	79	4	1	1	79	79.54
2014	84	5	3	9	252	83.62
2015	88	4	5	25	440	87.7
合 计	465	—	0	70	143	465

依据表 7.20 中的资料，直线趋势模型的 a 、 b 两参数值为：

$$\begin{cases} a = \frac{\sum y}{n} = \frac{465}{6} = 77.5 \\ b = \frac{\sum ty}{\sum t^2} = \frac{143}{70} = 2.04 \end{cases}$$

将 a 、 b 两值代入直线趋势模型，得

$$y_c=77.5+2.04t$$

预测该地区 2016 年甲产品可能达到的销售量（这时 $t=7$ ）为：

$$y_c = 77.5 + 2.04 \times 7 = 91.78 \text{ (千台)}$$

可见，用简捷法计算的各年趋势值和用一般计算方法计算的各年趋势值结果相同，以此预测的 2016 年该地区甲产品可能达到的销售量的结果也相同。

（2）抛物线方程。如果现象的发展，其逐期增长量的增长量（各期的二级增长量）大体相同，则可考虑曲线趋势——配合抛物线方程。抛物线的一般方程为：

$$y_c = a + bt + ct^2$$

上述抛物线方程式中，有 a 、 b 、 c 三个未定参数，根据最小平方方法的要求，同样用求偏导数的方法，导出以下由三个方程组成的联立方程组：

$$\begin{cases} \sum y = na + b \sum t + c \sum t^2 \\ \sum ty = a \sum t + b \sum t^2 + c \sum t^3 \\ \sum t^2 y = a \sum t^2 + b \sum t^3 + c \sum t^4 \end{cases}$$

同样，为了计算方便，我们可以通过假设 t ，使 $\sum t = 0$ ， $\sum t^3 = 0$ ，则上列联立方程组可简化为：

$$\begin{cases} \sum y = na + c \sum t^2 \\ \sum ty = b \sum t^2 \\ \sum t^2 y = a \sum t^2 + c \sum t^4 \end{cases}$$

将上述标准联立方程组求解，就可得到 a 、 b 、 c 之值，再将这三个参数代入抛物线方程，即为所求的抛物线方程模型。它同样可以帮助我们认识事物发展变化的规律，预测事物的未来。

（3）指数曲线方程。如果现象的发展，其环比发展速度或环比增长速度大体相同，则可考虑用曲线趋势——配合指数曲线方程。指数曲线的一般方程为：

$$y_c = ab^t$$

式中： a ——动态数列的基期水平；

b ——现象的一般发展速度；

t ——动态数列的时间。

a 、 b 为未定参数。

进行指数曲线拟合时，一般是将指数方程通过取对数转化成直线方程，然后按直线方程办法确定参数，再对直线方程求得的结果查反对数表还原。

先对上述方程式两边各取对数，得：

$$\lg y_c = \lg a + t \lg b$$

设: $Y=\lg y_c$, $A=\lg a$, $B=\lg b$

则: $Y=A+Bt$

根据最小平方原理, 推导出标准方程组为:

$$\begin{cases} \sum Y = nA + B \sum t \\ \sum tY = A \sum t + B \sum t^2 \end{cases}$$

同样设 t , 使 $\sum t = 0$, 则此联立方程组可简化为:

$$\begin{cases} \sum Y = nA \\ \sum tY = B \sum t^2 \end{cases}$$

将上述联立方程组求解, 即可求出 A 和 B 的值。由于 A 和 B 为对数值, 需要查反对数表求得 a 和 b 值。把参数代入指数曲线模型, 可以求得动态数列中与各观察值 (实际值) 对应的趋势值, 由此形成一条趋势线。

7.4.2 季节变动分析预测

在一个动态数列中, 除存在长期趋势外, 往往还存在季节变动。例如, 夏天汗衫、背心、冷饮的销售量就高于其他季节; 冬天围巾、皮衣、取暖器的销售量就比较大; 铁路客运量以春节前后为高峰。在另外一些情况下, 季节变动会引起设备和劳动力使用的不平衡、原料供应不足、运输量不够, 这给生产和人们生活带来某些影响。我们研究季节变动的目的, 主要是为了认识它、掌握它, 合理利用季节变动的规律, 以便为高效率组织工作、安排人们经济生活提供资料。

测定季节变动的方法很多, 从其是否考虑受长期趋势的影响来看, 有两种方法: 一种是不考虑长期趋势的影响, 直接根据原始的时间数列来计算, 常用的方法是按月 (季) 平均法; 另一种是根据剔除长期趋势影响后的数列资料来计算, 常用的方法是移动平均趋势剔除法。不管使用哪种方法来计算季节变动, 都需要用 3 年或更多年份的分月或分季资料作为基本数据进行计算分析, 这样才能较好地消除偶然因素的影响, 使季节变动的规律性更切合实际。

1. 按月 (季) 平均法

按月平均法亦称按季平均法。若是月资料就是按月平均; 若是季资料则是按季平均。其计算的一般步骤如下:

- (1) 列表。将各年同月 (季) 的数值列在同一栏内。
- (2) 将各年同月 (季) 数值加总, 并求出月 (季) 平均数。
- (3) 将所有月 (季) 数值加总, 求出总的月 (季) 平均数。
- (4) 求季节比率 (或季节指数), 其计算公式为:

$$\text{季节比率}=\frac{\text{各月平均数}}{\text{全期各月平均数}}\times100\% \tag{7-34}$$

如果计算出来的某月（季）的季节比率大于 100%，则表明现象在该月（季）为旺季；若某月（季）的季节比率小于 100%，则为淡季；如果等于 100%，则表明该月不受季节变动的影响。

【实例 7.23】 某商业企业毛线销售情况如表 7.21 所示，计算季节比率。

解：首先，计算月（季）平均数。如一月份为(400+400+460)÷3=420，其余月份类推。

其次，计算总的月（季）平均数。把 3 年 36 个月的资料全部相加后平均，或用 12 个月平均数求总平均数，即(222.66+227.84+255.00)÷3=235.17。

然后，计算各月平均数与总平均数的对比值，得出各月季节比率。如一月份为 420÷235.17=178.6%，其余月份类推。

表 7.21 某商业企业毛线销售季节变动表 单位：千克

月 份	第一年	第二年	第三年	三年平均	季节比率（%）
1	400	400	460	420	178.6
2	420	420	480	440	187.1
3	300	260	340	300	127.6
4	180	160	200	180	76.5
5	140	160	180	160	68.0
6	120	140	100	120	51.0
7	100	98	120	106	45.1
8	82	76	100	86	36.6
9	160	180	200	180	76.5
10	170	180	220	190	80.8
11	220	240	200	220	93.6
12	380	420	460	420	178.6
合 计	2 672	2 734	3 060	2 822	1 200.0
平 均	222.66	227.84	255.00	235.17	100.0

计算结果表明，毛线销售呈明显的季节变动。每年 12 月至第二年 3 月为销售旺季，其中 2 月份季节比率高达 187.1%而进入顶峰；从 4 月份开始进入淡季，一直延续到 11 月份，其中 8 月份季节比率仅有 36.6%而降低低谷。据此，我们可以预测未来的季节变动，组织好毛线的经营。

按月平均预测法的计算公式为：

$$\text{各月（季）预测值}=\text{上年各月（季）平均值}\times\text{各月（季）的季节比率}$$

假定第四年该企业的毛线销售量为 4 200 千克，则

$$\text{2 月份毛线销售的预测值为：}(3\,000\div12)\times187.1\%=446.51\text{（千克）}$$

8 月份毛线销售的预测值为： $(3\,000 \div 12) \times 36.6\% = 91.5$ （千克）

按月（季）平均法的优点是计算简便、容易理解，而缺点是不够精确，因为它没有消除时间数列中长期趋势的影响。要解决这一问题，就要用另一种方法，即趋势剔除法。

2. 移动平均趋势剔除法

移动平均趋势剔除法就是按移动平均法来剔除长期趋势的影响，再计算季节变动的方法。

【实例 7.24】 某公司连续 16 个季度的商品销售量资料如表 7.22 所示，采用趋势剔除法计算其季节比率。

表 7.22 某公司商品销售量资料及计算表 单位：万件

年 份	季 度	销售量	四项移动平均	正位移动平均趋势值 y_c	剔除趋势值 y/y_c
2011	3	15	—	—	—
	4	18	12.0	—	—
2012	1	6	12.5	12.25	48.98
	2	9	13.0	12.75	70.59
	3	17	13.5	13.25	128.30
	4	20	14.0	13.75	145.45
2013	1	8	14.5	14.25	56.14
	2	11	15.0	14.75	74.58
	3	19	15.5	15.25	124.59
	4	22	16.0	15.75	135.21
2014	1	10	16.5	16.25	61.54
	2	13	17.0	16.75	77.61
	3	21	17.5	17.25	121.74
	4	24	18.0	17.75	135.21
2015	1	12	—	—	—
	2	13	—	—	—

解：首先，采用四项（因为是季度资料）移动平均法，求出移动平均趋势值 y_c ，如表 7.22 所示。由于采用四项移动，其平均值对正两季中间，一般无法直接比较，所以应再进行一次两项移动平均，就得出对准各季的移正平均值，即表 7.22 中最后一栏的趋势值 y_c 。

其次，从原数列剔除已测定的长期趋势变动。剔除趋势值的方法有两种，一种是从实际观测值 y 减去趋势值 y_c ；另一种是实际观测值 y 除以趋势值 y_c 。我们以第二种为例列表计算，如表 7.22 所示。

原数列是由 16 个季度资料构成的，因采用 4 季移动平均法，所以由趋势值形成的新数列只有全时期中间的 12 个季度的数字，即表 7.22 中 2012—2014 年的 12 个季度的数字。

然后，根据剔除趋势值所得的结果分别计算季节比率。计算季节比率的主要依据就是各年各季的 y/y_c 的百分比数字。将表 7.22 中各年间季的百分比数字重新排成一列，分别求出各季的

平均数，列表计算结果如表 7.23 所示。

表 7.23 趋势剔除法季节比率计算表

单位：%

年 份	第一季	第二季	第三季	第四季	合 计
2012	48.98	70.59	128.30	145.45	393.32
2013	56.14	74.58	124.59	139.68	394.99
2014	61.54	77.61	121.74	135.21	396.10
合 计	166.66	222.78	374.63	420.34	1 184.41
平 均 数	55.55	74.26	124.88	140.11	394.8
季节比率	56.28	75.24	126.52	141.96	400.00

最后，由于季节比率是围绕 100%上下波动的，按四个季度计算的季节比率之和应为 400%（按月的季节比率之和应为 1 200%）；否则就应计算调整系数设法调整。调整系数的计算公式为：

$$\text{调整系数} = 400\% (1\ 200\%) \div \text{调整前各季（月）季节比率之和}$$

$$\text{调整后的季节比率} = \text{调整前各季（月）季节比率} \times \text{调整系数}$$

在本例中，由于各季百分比数值的平均数之和为 394.80%，因此需要校正。调整系数=400÷394.80=1.013 17。以这一系数分别乘各季的平均数，即得表 7.23 中的季节比率。例如，第一季度调整后的季节比率为：55.55%×1.013 17=56.28%；其余季度以此类推。

季节比率说明现象在剔除了长期趋势影响后，一年中各季的波动程度，由于是以百分比表示的典型波动，所以也可以用于季节比率的预测。

必须指出，某些现象的季节变动并非是永恒的规律，随着科技的进步和人们生活习惯的改变，某些社会现象的季节变动会被削弱甚至完全消失。例如，鸡在冬天通常不下蛋，但在现代化的养鸡场，用电灯代替阳光延长白昼，室内设有空调，饲料又有专门的配方，讲究营养，鸡在冬天照样下蛋。

在实际中，季节变动的分析已被推广为凡在短期内，现象有周期性的规律变动，都可称为季节变动，也可用这类方法去测定。如一周内，哪天公园的游客最多，超市的销售额在星期几最高；一天内，哪些时间（小时）交通最拥挤，什么时段商场客流量最大，等等。

7.5 利用Excel进行时间数列分析

7.5.1 利用Excel进行水平分析与速度分析

对时间数列的水平分析，主要是计算动态平均数（也称为序时平均数、平均发展水平）、增长量与平均增长量。对时间数列的速度分析主要是计算发展速度、平均发展速度及增长速度与平均增长速度。以【实例 7.25】和【实例 7.26】为例，分别说明利用 Excel 进行时间数列水平分析和速度分析的计算方法。

1. 时间数列的水平分析

【实例 7.25】 根据我国 2000~2010 年的粮食产量的数据（资料来源：国家统计局，单位：万吨），计算粮食产量的动态平均数、粮食产量逐年增长量、累计增长量和平均增长量。

解：具体操作步骤如下，结果如图 7.1 所示。

	A	B	C	D	E	F	G
1	年份	粮食产量（万吨）	逐期增长量（万吨）	累计增长量（万吨）			
2	2000	46 251	—	—			
3	2001	45 264	-987	-987			
4	2002	45 706	442	-545			
5	2003	43 070	-2 636	-3 181			
6	2004	46 947	3 877	696	粮食产量的动态平均数	平均增长量	
7	2005	48 402	1 455	2 151	48 745.27		839
8	2006	49 804	1 402	3 553			
9	2007	50 160	356	3 909			
10	2008	52 871	2 711	6 620			
11	2009	53 082	211	6 831			
12	2010	54 641	1 559	8 390			

图 7.1 我国 2000~2010 年粮食产量水平分析计算结果

首先，在 A 列输入“年份”，在 B 列输入“粮食产量”。

第一步：计算粮食产量的动态平均数。可直接利用函数 AVERAGE 计算，单击任一空单元格，本例为 E7，输入“=AVERAGE(B2:B12)”得结果 48 745.27 万吨。

第二步：计算逐期增长量。在 C3 单元格中输入“=B3-B2”，回车得结果；然后使用填充柄功能按住鼠标左键向下拖，至 C12 单元格，放开鼠标，可得 C4~C12 结果，即 2002~2010 年的逐期增长量自动填入 C4~C12 单元格。

第三步：计算累计增长量。在 D3 单元格中输入“=B3-46251”，回车得结果；然后使用填充柄功能按住鼠标左键向下拖，至 D12 单元格，放开鼠标，可得 D4~D12 结果，即 2002~2010 年的累积增长量自动填入 D4~D12 单元格。

第四步：计算平均增长量。在任一空单元格，本例为 G7，输入“=D12/(11-1)”或“=SUM(C3:C12)/(11-1)”，按回车键，即可得到平均增长量 839 万吨。

2. 时间数列的速度分析

【实例 7.26】 根据我国 2000~2010 年的粮食产量的数据（资料来源：国家统计局，单位：万吨），计算粮食产量的发展速度、增长速度。

解：具体操作步骤如下，结果如图 7.2 所示。

	A	B	C	D	E	F
1	年份	粮食产量（万吨）	环比发展速度（%）	定基发展速度（%）	环比增长速度（%）	定基增长速度（%）
2	2000	46 251	—	100.00	—	0.00
3	2001	45 264	97.87	97.87	-2.13	-2.13
4	2002	45 706	100.98	98.82	0.98	-1.18
5	2003	43 070	94.23	93.12	-5.77	-6.88
6	2004	46 947	109	101.5	9	1.5
7	2005	48 402	103.1	104.65	3.1	4.65
8	2006	49 804	102.9	107.68	2.9	7.68
9	2007	50 160	100.71	108.45	0.71	8.45
10	2008	52 871	105.4	114.31	5.4	14.31
11	2009	53 082	100.4	114.77	0.4	14.77
12	2010	54 641	102.94	118.14	2.94	18.14

图 7.2 我国 2000~2010 年粮食产量速度分析计算结果

第一步：计算粮食产量的环比发展速度。在 C3 单元格中输入“=ROUND(B3/B2*100,2)”，回车得结果；然后使用填充柄功能按住鼠标左键向下拖，至 C12 单元格，放开鼠标，可得 C4~C12 结果。

第二步：计算粮食产量定基发展速度。在 D3 单元格中输入“=ROUND(B3/46251*100,2)”，回车得结果；然后使用填充柄功能按住鼠标左键向下拖，至 D12 单元格，放开鼠标，可得 D4~D12 结果。

第三步：计算粮食产量的环比增长速度。在 E3 单元格中输入“=C3-100”，回车得结果；然后使用填充柄功能按住鼠标左键向下拖，至 E12 单元格，放开鼠标，可得 E4~E12 结果。

第四步：计算粮食产量的定基增长速度。在 F3 单元格中输入“=D3-100”，回车得结果；然后使用填充柄功能按住鼠标左键向下拖，至 F12 单元格，放开鼠标，可得 F4~F12 结果。

7.5.2 利用Excel进行长期趋势分析

以直线趋势为例说明长期趋势的测定与预测的方法。测定直线趋势的方法主要有移动平均法和最小二乘法。

1. 移动平均法

移动平均法按一定的间隔逐期移动，计算一系列动态平均数，从而形成一个由动态平均数组成的新的时间数列，修匀原时间数列，显示出长期趋势。以【实例 7.27】说明移动平均法的操作方法。

在 Excel 中，使用移动平均法测定长期趋势，可以利用公式与 AVERAGE 函数，也可利用 Excel 提供的“移动平均”工具进行分析。由于公式与函数方法前面已多次讲过，而且只能获得数据，不能直接获得长期趋势图，因此长期趋势主要以移动平均工具来分析，公式与函数计算

移动平均的应用在季节变动分析一节再详细讲解。

【实例 7.27】 根据我国 2000~2010 年的粮食产量的数据（资料来源：国家统计局，单位：万吨），用移动平均法分析我国粮食产量的长期发展趋势。

解：具体操作步骤如下。

第一步：单击“工具”菜单，选择“数据分析”选项。打开“数据分析”对话框，从其对话框的“分析工具”列表中选择“移动平均”，单击“确定”按钮，打开“移动平均”对话框。

第二步：确定输入区域和输出区域。

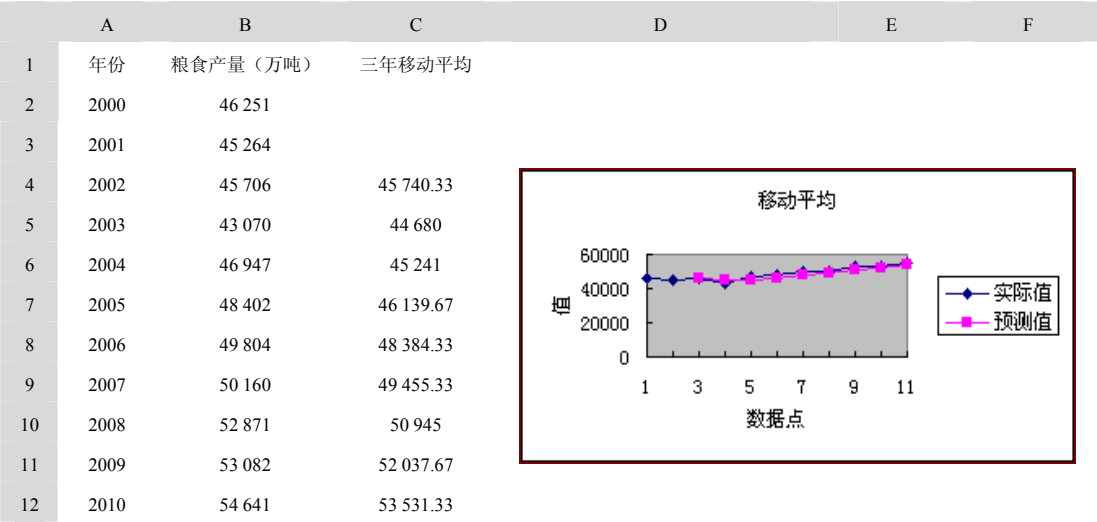
确定输入区域。在“输入区域”框中输入分析数据所在的单元格区域，在本例中，输入区域为\$B\$2:\$B\$12。“间隔”框中输入移动平均的项数，本例中为 3。

确定输出选项。在“输出选项”对话框中可以指定结果的输出去向，输出去向有三种。在“输出区域”框中输入输出结果所在的单元格区域的左上角的单元格的行列号，在本例中，输出区域为\$C\$2。也可以通过选择“新工作表组”或“新工作簿”将结果放在新工作表或新工作簿。

若选中“图表输出”，则显示移动平均统计图。本例中为选中。

若选中“标准误差”，则输出移动平均值与原数据的标准差。

第三步：单击“确定”按钮后，在指定位置给出移动平均计算结果，如图 7.3 所示。



2. 最小二乘法

为了用最小二乘法进行时间数列回归分析,Excel 不仅提供了 LINEST、INTERCEPT、SLOPE 等函数，而且提供了一个专门的数据分析工具——回归分析工具。回归分析工具在后面章节中详细讲解，这里只进行简单介绍。利用最小二乘法测定直线趋势，就要为时间数列配合一条趋势直线，以时间为自变量，分析动态指标随时间的变动而变动的规律。在此主要说明使用回归

分析工具进行直线趋势测定的操作方法。

【实例 7.28】 根据我国 2000~2010 年的粮食产量的数据（资料来源：国家统计局，单位：万吨），用最小二乘法分析我国粮食产量的长期发展趋势。

解：具体操作步骤如下。

第一步：输入“年份”和“粮食产量”，并且为年份排出年序（ t ）。

第二步：单击“工具”菜单，选择“数据分析”选项。打开“数据分析”对话框，从其对话框的“分析工具”列表中选择“回归”，单击“确定”按钮，打开“回归”对话框。

第三步：确定输入区域和输出区域。

确定输入区域。在“Y 值输入区域”框中输入因变量数据所在的单元格区域，在本例中，输入区域为\$C\$2:\$C\$12。在“X 值输入区域”框中输入自变量数据所在的单元格区域，在本例中，输入区域为\$B\$2:\$B\$12。

确定输出选项。在“输出选项”对话框中可以指定结果的输出去向，输出去向有三种。在“输出区域”框中输入输出结果所在的单元格区域的左上角的单元格的行列号，在本例中，输出区域为\$F\$2。也可以通过选择“新工作表组”或“新工作簿”将结果放在新工作表或新工作簿。

一般情况下，要求给出趋势值和趋势线图，则要选中“残差”与“线形拟合图”复选框。

第四步：将残差表中的预测值复制到 D 列，以便于观察，如图 7.4 所示。

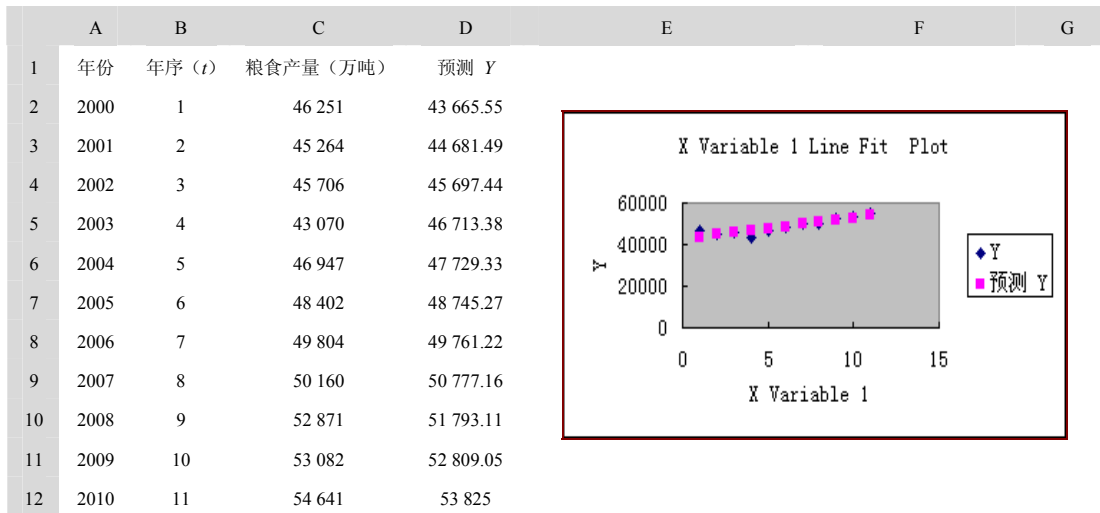


图 7.4 最小二乘法计算结果

7.5.3 利用Excel进行季节变动分析

季节变动是指受自然季节和社会习俗等因素影响而发生的有规律的、周期性的变动。测定季节变动的方法有两种：按月（季）平均法和移动平均趋势剔除法。

1. 按月平均法

【实例 7.29】 某啤酒厂近五年全年及分月啤酒销售量数据如表 7.24 所示，结合五年分月数据，利用 Excel 用按月平均法测定季节变动。

表 7.24 某啤酒厂近五年全年及分月啤酒销售量

年 份	啤酒销售量											
	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
第一年	18	10	4	4	11	15	18	12	10	25	30	31
第二年	20	12	5	6	25	30	42	21	15	40	72	58
第三年	27	18	10	9	40	55	90	25	17	75	80	72
第四年	40	30	18	15	45	80	114	40	35	90	105	73
第五年	48	36	23	30	30	97	125	47	45	103	128	96

解：按月平均法具体操作步骤如下，各步计算数据如图 7.5 所示。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	啤酒销售量													
2	年份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月	合计
3	第一年	18	10	4	4	11	15	18	12	10	25	30	31	188
4	第二年	20	12	5	6	25	30	42	21	15	40	72	58	346
5	第三年	27	18	10	9	40	55	90	25	17	75	80	72	518
6	第四年	40	30	18	15	45	80	114	40	35	90	105	73	685
7	第五年	48	36	23	30	30	97	125	47	45	103	128	96	808
8	合计	153	106	60	64	151	277	389	145	122	333	415	330	2 545
9	同月平均	30.6	21.2	12	12.8	30.2	55.4	77.8	29	24.4	66.6	83	66	42.42
10	季节比率%	72.1	49.98	28.29	30.17	71.19	130.6	183.4	68.36	57.5	157.0	195.7	155.6	1 200

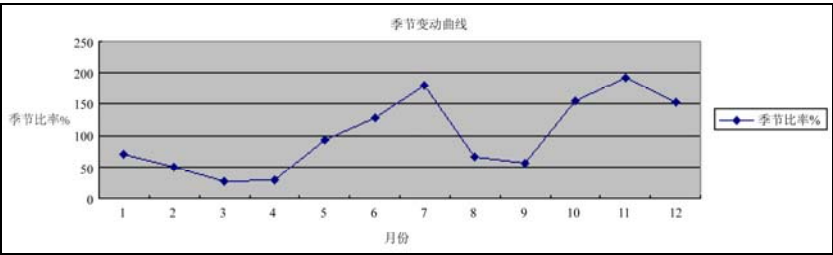


图 7.5 按月平均法分析季节变动数据图

第一步：按已知数据资料列出计算表，将各年同月的数值列在同一列内。

第二步：计算各年合计与各年同月数值之和。计算每年的啤酒销量总数：单击 N3 单元格，输入“=SUM(B3:M3)”，回车得结果，然后使用填充柄功能按住鼠标左键向下拖，至 N7 单元格，放开鼠标，可得 N4~N7 结果。计算各年同月销售总数：单击 B8 单元格，输入“=SUM(B3:B7)”，回车得结果；然后使用填充柄功能按住鼠标左键向右拖，至 N8 单元格，放开鼠标，可得 C8~N8 结果。

第三步：计算同月平均数与总的月平均数。计算同月平均数：单击 B9 单元格，输入“=

B8/5”，回车得1月的月平均数；然后使用填充柄功能按住鼠标左键向右拖，至M9单元格，放开鼠标，可得C9~M9结果。计算总的月平均数：单击N9单元格，输入“=N8/60”，回车得结果为42.4167。

第四步：计算季节比率。单击B10单元格，输入“=B9*100/43.2167”，回车得1月的季节比率；然后使用填充柄功能按住鼠标左键向右拖，至M10单元格，放开鼠标，可得C10~M10结果。

第五步：计算季节比率之和，绘制季节变动曲线，这一步根据需要可做可不做。单击N10单元格，输入“=SUM(B10:M10)”，回车得季节比率之和为1200（如果是季度资料，则季节比率之和为400）。根据季节比率，可绘制季节变动曲线，绘图方法参见本书统计图的绘制。

2.移动平均趋势剔除法

直接用按月平均法忽略了长期趋势的影响，因此得出的季节比率不够精确。为了弥补这一缺点，可以采用移动平均趋势剔除法来测定季节变动。移动平均趋势剔除法分析季节变动有两种方法：乘法型时间数列季节变动分析和加法型时间数列季节变动分析。这里以乘法型时间数列季节变动分析说明移动趋势剔除法的操作方法。

【实例 7.30】 仍沿用【实例 7.29】的数据资料，本例中将其转化为季度资料，如表 7.25 所示，用移动平均趋势剔除法测定季节变动。

表 7.25 某啤酒厂近五年全年及各季度啤酒销售量

年 份	啤酒销售量			
	第 1 季度	第 2 季度	第 3 季度	第 4 季度
第一年	32	30	40	86
第二年	37	61	78	170
第三年	55	104	132	227
第四年	88	140	189	268
第五年	107	205	217	327

解：移动平均趋势剔除法具体操作步骤如下。

前四步计算数据如图 7.6 所示。

第一步：按已知数据资料列出计算表。

第二步：计算四个季度的移动平均数。计算移动平均数可以采用移动平均工具，也可以使用公式与函数。移动平均工具在上一小节中已讲解过，本例采用公式与函数方法来计算。单击 D4 单元格，输入“=AVERAGE(C2:C5)”，回车得结果；然后使用填充柄功能按住鼠标左键向下拖，至 D20 单元格，放开鼠标，可得 D5~D20 结果，即自动填入 D5~D20 的移动平均数。

第三步：移正平均。因为本例是偶数项移动平均，所以还需将第一次移动的平均值再进行两项“移正”平均。如果是奇数项移动平均，则该步骤省去。单击 E4 单元格，输入“=AVERAGE(D4:D5)”，回车得结果；然后使用填充柄功能按住鼠标左键向下拖，至 D19 单元格，放开鼠标，可得 E5~E19 结果，即自动填入 E5~E19 的移动平均数。

第四步：消除长期趋势。本例采用乘法模型，因此，将原数列除以趋势值以消除长期趋势。单击 F4 单元格，输入“= C4*100/E4”，回车得结果；然后使用填充柄功能按住鼠标左键向下拖，至 F19 单元格，放开鼠标，可得 F5~F19 结果，即自动填入 F5~F19 的剔除长期趋势后的数值。

	A	B	C	D	E	F
1	年份	季度	销售量（Y）	一次移动平均	二次移动平均（T）	Y/T*100
2	第一年	1	32			
3		2	30			
4		3	40	47	47.625	83.989 501
5	第二年	4	86	48.25	52.125	164.988 01
6		1	37	56	60.75	60.905 35
7		2	61	65.5	76	80.263 158
8	第三年	3	78	86.5	88.75	87.887 324
9		4	170	91	96.375	176.394 29
10		1	55	101.75	108.5	50.691 244
11	第四年	2	104	115.25	122.375	84.984 678
12		3	132	129.5	133.625	98.783 91
13		4	227	137.75	142.25	159.578 21
14	第五年	1	88	146.75	153.875	57.189 277
15		2	140	161	166.125	84.273 89
16		3	189	171.25	173.625	108.855 29
17	第六年	4	268	176	184.125	145.553 29
18		1	107	192.25	195.75	54.661 558
19		2	205	199.25	206.625	99.213 551
20	第七年	3	217	214		
21		4	327			

图 7.6 移动平均及剔除趋势数据

第五步：计算季节比率。用图 7.6 中“Y/T*100”得到的数据重新编排，得到图 7.7 中左边数据表中前五行的基本数据（本例将 7.6 图中 Y/T*100 的数据四舍五入，保留两位小数）。然后利用按月平均法计算季节比率，具体步骤参见按月平均法，本例省略。最终结果如图 7.7 所示。

	A	B	C	D	E	F	G	H	I	J	K
1	啤酒销售量										
2	年份	第 1 季度	第 2 季度	第 3 季度	第 4 季度	合计					
3	第一年			83.99	164.99						
4	第二年	60.91	80.26	87.89	176.39						
5	第三年	50.69	84.98	98.78	159.58						
6	第四年	57.19	84.27	108.86	145.55						
7	第五年	54.66	99.21								
8	合计	223.45	348.72	379.52	646.51	1 598.2					
9	同季平均	55.862 5	87.18	94.88	161.627 5	99.887 5					
10	季节比率%	55.925 42	87.278 19	94.986 86	161.809 5	400					

季节变动曲线

图 7.7 利用移动平均趋势剔除法分析季节变动数据图

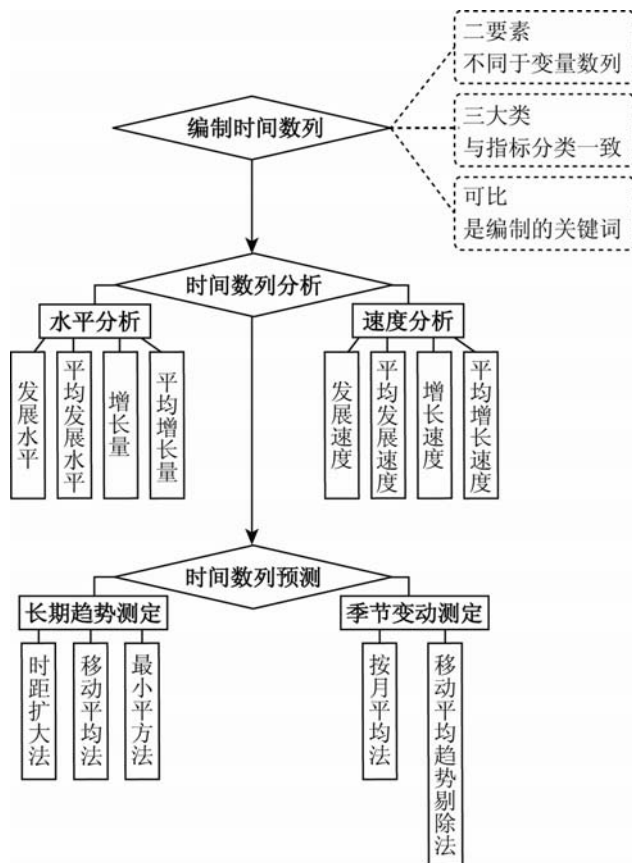


统计术语

时间数列	time-series	发展水平	time-series data
发展速度	rate of expansion	增长速度	rate of growth
长期趋势	trend	季节变动	seasonal effects
循环变动	cyclical effects	不规则变动	irregular fluctuations
最小平方法	least square method	移动平均	moving average
定基	fixed base		



重点知识梳理



习题与实践训练

一、判断题

1. 在各种时间数列中，指标值的大小都受到指标所反映的时期长短的制约。（ ）

2. 定基发展速度反映了现象在一定时期内发展的总速度, 环比发展速度反映了现象较前一年的增长程度。 ()
3. 若逐期增长量每年相等, 则其各年的环比发展速度是年年下降的。 ()
4. 环比速度与定基速度之间存在如下关系式: 各期环比增长速度的连乘积等于定基增长速度。 ()
5. 用水平法计算的平均发展速度只取决于最初发展水平和最末发展水平, 与中间各期发展水平无关。 ()
6. 呈直线趋势的时间数列, 其各期环比发展速度大致相同。 ()
7. 计算平均发展速度有两种方法, 即几何平均法和方程式法, 这两种方法是根据分析目的不同划分的。 ()
8. 平均发展速度是环比发展速度的平均数, 也是一种序时平均数。 ()
9. 将某班学生按考试成绩分组形成的数列是时点数列。 ()
10. 增长量与基期发展水平指标对比, 得到的是发展速度指标。 ()

二、单项选择题

1. 下列数列中哪一个属于时间数列 ()。
 - A. 学生按学习成绩分组形成的数列
 - B. 工业企业按地区分组形成的数列
 - C. 职工按工资水平高低排列形成的数列
 - D. 出口额按时间先后顺序排列形成的数列
2. 已知各期环比增长速度为 2%、5%、8% 和 7%, 则相应的定基增长速度的计算为 ()。
 - A. $(102\% \times 105\% \times 108\% \times 107\%) - 100\%$
 - B. $102\% \times 105\% \times 108\% \times 107\%$
 - C. $2\% \times 5\% \times 8\% \times 7\%$
 - D. $(2\% \times 5\% \times 8\% \times 7\%) - 100\%$
3. 已知某企业 1 月、2 月、3 月、4 月的平均职工人数分别为 190 人、195 人、193 人和 201 人。则该企业一季度的平均职工人数的计算方法为 ()。
 - A. $\frac{190 + 195 + 193 + 201}{4}$
 - B. $\frac{190 + 195 + 193}{3}$
 - C. $\frac{\frac{190}{2} + 195 + 193 + \frac{201}{2}}{4 - 1}$
 - D. $\frac{\frac{190}{2} + 195 + 193 + \frac{201}{2}}{4}$
4. 平均发展速度是 ()。
 - A. 定基发展速度的算术平均数
 - B. 环比发展速度的算术平均数
 - C. 环比发展速度的几何平均数
 - D. 增长速度加上 100%
5. 定基增长速度与环比增长速度的关系是 ()。
 - A. 定基增长速度是环比增长速度的连乘积
 - B. 定基增长速度是环比增长速度之和
 - C. 各环比增长速度加 1 后的连乘积减 1
 - D. 各环比增长速度减 1 后的连乘积减 1
6. 计算序时平均数时, “首末折半法” 适用于 ()。

- A. 时期数列计算序时平均数
 B. 间隔相等的间断时点数列计算序时平均数
 C. 间隔不等的间断时点数列计算序时平均数
 D. 由两个时点数列构成的相对数动态数列计算序时平均数
7. 间隔不等的时点数列求序时平均数的公式是（ ）。
- A.
$$\bar{a} = \frac{\left(\frac{a_1 + a_2}{2}\right)f_1 + \left(\frac{a_2 + a_3}{2}\right)f_2 + \cdots + \left(\frac{a_{n-1} + a_n}{2}\right)f_{n-1}}{\sum f}$$
- B.
$$\bar{a} = \frac{\sum a}{n}$$
- C.
$$\bar{a} = \frac{\sum af}{\sum f}$$
- D.
$$\bar{a} = \frac{\frac{a_1}{2} + a_2 + \cdots + \frac{a_n}{2}}{n-1}$$
8. 设 2011—2015 年各年的环比增长速度为 6%、7%、8%、9%和 10%，则平均增长速度为（ ）。
- A. $\sqrt[5]{6\% \times 7\% \times 8\% \times 9\% \times 10\%}$
 B. $\sqrt[5]{106\% \times 107\% \times 108\% \times 109\% \times 110\%}$
 C. $\sqrt[5]{106\% + 107\% + 108\% + 109\% + 110\%}$
 D. $\sqrt[5]{106\% \times 107\% \times 108\% \times 109\% \times 110\%} - 100\%$
9. 假定某产品产量 2015 年比 2010 年增加了 35%，则 2010—2015 年的平均发展速度为（ ）。
- A. $\sqrt[5]{35\%}$ B. $\sqrt[5]{135\%}$ C. $\sqrt[5]{35\%}$ D. $\sqrt[5]{135\%}$
10. 已知某企业 1 月、2 月、3 月、4 月的月初职工人数分别为 190 人、195 人、193 人和 201 人。则该企业一季度的平均职工人数的计算方法为（ ）。
- A. $\frac{190 + 195 + 193 + 201}{4}$ B. $\frac{190 + 195 + 193}{3}$
 C. $\frac{\frac{190}{2} + 195 + 193 + \frac{201}{2}}{4-1}$ D. $\frac{\frac{190}{2} + 195 + 193 + \frac{201}{2}}{4}$

三、多项选择题

1. 一个时间数列的基本要素包括：（ ）。
- A. 变量 B. 次数
 C. 现象所属的时间 D. 现象所属的地点
 E. 反映现象的统计指标值
2. 按统计指标表现的形式看，时间数列可分为如下几种（ ）。
- A. 总量指标时间数列 B. 相对指标时间数列
 C. 平均指标时间数列 D. 时期指标时间数列 E. 时点指标数列

3. 下面哪几项是时期数列 ()。
 - A. 我国近年来的耕地总面积
 - B. 我国历年新增人口数
 - C. 我国历年图书出版量
 - D. 我国历年黄金储备
 - E. 某地区国有企业历年资金利税率
4. 定基发展速度和环比发展速度的关系是 ()。
 - A. 两者都属于速度指标
 - B. 相应环比发展速度的连乘积等于定基发展速度
 - C. 相应定基发展速度的连乘积等于环比发展速度
 - D. 相邻两个定基发展速度之商等于相应的环比发展速度
 - E. 相邻两个环比发展速度之商等于相应的定基发展速度
5. 累积增长量与逐期增长量 ()。
 - A. 前者基期不变, 后者基期逐期在变动
 - B. 二者存在关系式: 逐期增长量之和=累积增长量
 - C. 相邻的两个逐期增长量之差等于相应的累积增长量
 - D. 根据这两个增长量都可以计算平均每期增长量
 - E. 这两个增长量都属于速度分析指标
6. 下列哪些指标是序时平均数 ()。
 - A. 一季度平均每月的职工人数
 - B. 某产品产量某年各月的平均增长量
 - C. 某企业职工第四季度人均产值
 - D. 某商场职工某年月平均人均销售额
 - E. 某地区近几年出口商品贸易额平均增长速度
7. 增长 1% 的绝对值 ()。
 - A. 等于前期水平除以 100
 - B. 等于逐期增长量除以环比增长速度
 - C. 等于逐期增长量除以环比发展速度
 - D. 表示增加一个百分点所增加的绝对量
 - E. 表示增加一个百分点所增加的相对量
8. 下面属于时点数列的有 ()。
 - A. 历年旅客周转量
 - B. 某工厂每年设备台数
 - C. 历年商品销售量
 - D. 历年牲畜存栏数
 - E. 某银行储户存款余额
9. 时期数列的特点有 ()。
 - A. 数列中各个指标数值不能相加

- B. 数列中各个指标数值可以相加
 - C. 数列中每个指标数值大小与其时间长短无直接关系
 - D. 数列中每个指标数值的大小与其时间长短有直接关系
 - E. 数列中每个指标数值，通常是通过连续不断登记而取得的
10. 常用的长期趋势时间数列修匀的方法主要有（ ）。
- A. 时距扩大法 B. 几何平均法 C. 最小平方法
 - D. 移动平均法 E. 动态平均法

四、填空题

1. 时间数列是由反映_____变化和_____变化的两个数列所构成的。
2. 总量指标时间数列按反映经济现象性质的不同可分为_____和_____两种。
3. 各年末商品库存量数列属于_____数列，各年的基建投资额数列属于_____数列。
4. 社会经济现象发展的动态分析主要包括_____和_____两部分。
5. 静态平均数是根据_____计算的，序时平均数则是根据_____计算的。
6. 某高校在校学生 2013 年比 2012 年增加 8%，2014 年比 2013 年增加 12%，2015 年比 2014 年增加 10%，则这三年共增加学生_____。
7. 各项环比发展速度的_____等于定基发展速度，累计增长量等于各项逐期增长量的_____。
8. 本期定基发展速度与前期定期发展速度之比等于_____，本期累计增长量与本期逐期增长量之差等于_____。
9. 某公司生产某产品，一月份生产 2 000 件，废品率为 0.9%；二月份生产 2 200 件，废品率为 1.1%；三月份生产 1 900 件，废品率为 0.8%；则第一季度该公司产品的平均废品率为_____。
10. 计算季节比率通常有_____和_____两种方法。

五、应用能力训练题

1. 新华小区自行车库 4 月 1 日存自行车 320 辆，4 月 6 日调出 70 辆，4 月 18 日进货 120 辆，4 月 26 日调出 80 辆，直至月末自行车数量未发生变动。问该库 4 月份平均库存自行车多少辆？
2. 我国餐饮业年末从业人数资料如下表所示。

年 份	2009	2010	2011	2012	2013
年末从业人数（万人）	200.6	220.3	227.8	243.7	246.8

计算 2009—2013 年我国餐饮业平均从业人数。

3. 某公司 2015 年下半年各月末工人数及其比重资料如下。

月 份	6	7	8	9	10	11	12
月末工人数（人）	550	580	560	565	600	590	590
工人占全部职工人数的比重（%）	80.0	86.0	81.0	80.0	90.0	87.0	85.0

计算该公司 2015 年下半年工人占全部职工人数的平均比重。

4. 某商店商品销售额及销售人员人数资料如下。

月 份	3	4	5	6
商品销售额（万元）	165.0	198.0	177.0	216.9
月末销售人员人数（人）	200	205	208	206

根据上表资料计算：

- （1）第二季度该店平均每月商品销售额。
- （2）第二季度平均售货员人数。
- （3）第二季度平均每售货员的销售额。
- （4）4、5、6 各月份（分别）的平均每个售货员的销售额。
- （5）第二季度平均每月每个售货员的销售额。

5. 某地区某种产品产量资料如下。

年 份	产量（万台）	累计增长量（万台）	定基发展速度（%）	环比发展速度（%）
2010	200			
2011		50		
2012			125	
2013				115
2014				125
2015		160		

根据上表资料计算：

- （1）表中空格所缺统计指标。
- （2）平均增长量指标。
- （3）平均发展速度指标。
- （4）平均增长速度指标。

6. 已知 2014 年我国国内生产总值（GDP）为 636 463 亿元，若以平均每年增长 6% 的速度发展，试计算到 2020 年将达到什么水平？

7. 某地区 2015 年第三产业增加值为 100 亿元，如果以后每年增长 6.5%，试问多少年后才能达到 200 亿元？

8. 某市大型电器商场 2009—2015 年销售某家用电器资料如下。

年 份	2009	2010	2011	2012	2013	2014	2015
销售额（百万元）	5.6	11.4	17.1	24	31.3	38	43.9

试计算：

- （1）逐期增长量和累计增长量。
- （2）定基发展速度和环比发展速度。
- （3）定基增长速度和环比增长速度。
- （4）增长 1% 的绝对值。

(5) 平均增长速度。

(6) 用最小平方法求出直线趋势方程，并预测 2016 年销售额。

9. 请从时间长短、起伏规律和形成原因等三个方面判断下面这些现象属于时间数列构成因素中的哪一个？

(1) 超市的顾客人数，周末达到高峰。

(2) 银行的活期储蓄额，发放工资前减少，发放工资后增多。

(3) 我国的进出口贸易额从长时间来看是不断增长的。

(4) 由于媒体对“苏丹红”的报道，造成部分副食品需求量的急剧下降。

(5) 耐用消费品如电视、冰箱周期性更新导致需求量变化。

(6) 铁路部门运送旅客的客运量，在一年中有几个时段为高峰，另外几个时段为低谷。

(7) 玩具的销售量每年在 12 月份达到最大额。

(8) 据中国互联网络中心（CNNIC）调查资料，我国网民上网高峰集中在 20 点和 21 点，最高峰是 20 点，有超过 60% 的网民经常在这一时点上网。次高峰在 14 点，低谷则是凌晨到早上 8 点。

10. 某地区甲产品 2012—2015 年各季收购量统计资料如下表所示。

某地区甲产品收购量统计资料

单位：万吨

年 份	一季度	二季度	三季度	四季度
2012	15	7	10	20
2013	16	8	12	20
2014	18	10	14	24
2015	21	17	19	27

根据上表资料：

(1) 用移动平均法对该动态数列进行修匀。

(2) 用按月（季）平均法计算其季节比率。

(3) 假设 2016 年全年计划收购量为 120 万吨，则将计划目标分解，每季度计划收购量应分别为多少？

11. 我国 2000—2014 年城镇居民家庭人均可支配收入资料如下。

年 份	城镇居民家庭人均可支配收入（元）
2000	6 280.0
2001	6 859.6
2002	7 702.8
2003	8 472.2
2004	9 421.6
2005	10 493.0
2006	11 759.0
2007	13 786.0
2008	15 781.0
2009	17 175.0

续表

年 份	城镇居民家庭人均可支配收入（元）
2010	19 109.0
2011	21 809.8
2012	24 564.7
2013	26 955.1
2014	28 844.0

请利用 Excel 软件完成下列计算：

（1）编制 2000—2014 年城镇居民家庭人均可支配收入逐期增长量、累积增长量数列并计算动态平均数和平均增长量。

（2）编制 2000—2014 年城镇居民家庭人均可支配收入的环比发展速度、定基发展速度、环比增长速度、定基增长速度数列，并计算平均发展速度和平均增长速度。

（3）分别用移动平均法和最小平方方法进行长期趋势测定。

12. 某地区 2012—2015 年各季度羊绒衫销售量的资料如下。

单位：百件

季 度	2012	2013	2014	2015
1	450	500	580	700
2	150	180	200	250
3	250	280	320	380
4	500	560	630	750

请利用 Excel 软件完成下列计算：

（1）利用按季平均法测定季节比率。

（2）利用移动平均趋势剔除法测定季节比率。

13. 2015 年香港政府统计处综合住户统计调查发布了从 1994—2014 年相关就业人员每月就业收入中位数（港元），具体资料如下。

年 份	每月就业收入中位数（港元）	20~24 岁具专上教育程度的就业人士每月就业收入中位数（港元）
1994	8 000	9 000
1995	9 000	10 000
1996	10 000	10 000
1997	10 000	11 000
1998	10 000	11 000
1999	10 000	10 000
2000	10 000	10 000
2001	11 000	10 000
2002	10 400	8 500
2003	10 000	8 000
2004	10 000	7 500

续表

年 份	每月就业收入中位数（港元）	20~24 岁具专上教育程度的就业人士每月就业收入中位数（港元）
2005	10 000	8 000
2006	10 500	8 000
2007	11 000	8 500
2008	11 000	9 000
2009	11 000	8 500
2010	11 800	8 700
2011	12 000	9 100
2012	13 000	10 000
2013	13 500	10 000
2014	15 000	10 800

（资料来自香港政府统计处综合住户统计调查，以上数据均不包括外籍家庭护工。）

请思考，根据这两个时间数列，利用到目前为止我们所学到的统计知识，你可以从中进行怎样的分析，获得哪些信息，从而去认识所研究总体的特征。



本章案例

金砖国家联合统计手册 2015 ——2015 国民经济核算概况（2000—2014 年）

指标	2000	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
GDP（现价，亿美元）											
巴西	6 570	8 920	11 080	13 960	16 950	16 660	22 090	26 130	24 120	23 910	23 460
俄罗斯	2 600	7 640	9 900	13 000	16 610	12 230	15 249	19 067	20 010	20 802	18 810
印度	4 766	8 342	9 480	12 390	12 260	13 660	17 080	18 430	18 365	18 750	20 691
中国	12 053	22 693	27 303	35 247	45 608	50 597	60 404	74 956	84 614	94 946	103 610
南非	1 360	2 577	2 720	2 990	2 870	2 970	3 750	4 170	3 970	3 660	3 500
人均 GDP（现价，美元）											
巴西	3 788	4 817	5 912	7 367	8 846	8 610	11 298	13 235	12 104	11 892	11 571
俄罗斯	1 772	5 323	6 920	9 101	11 635	8 563	10 675	13 338	13 976	14 494	12 874
印度	468	754	845	1 089	1 063	1 168	1 440	1 511	1 486	1 499	1 633
中国	955	1 741	2 083	2 675	3 443	3 801	4 515	5 577	6 264	6 995	7 595
南非	3 034	5 442	5 664	6 146	5 818	5 938	7 389	8 090	7 592	6 890	6 483

指标	2000	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
GDP 增长率 (比上年增长, %)											
巴西	4.4	3.1	4.0	6.0	5.0	-0.2	7.6	3.9	1.8	2.7	0.1
俄罗斯	10.0	6.4	8.2	8.5	5.2	-7.8	4.5	4.3	3.4	1.3	0.6
印度	7.6	13.9	16.3	16.1	12.9	15.1	20.2	15.7	13.1	13.6	11.5
中国	8.4	11.3	12.7	14.2	9.6	9.2	10.6	9.5	7.7	7.7	7.4
南非	4.2	5.3	5.6	5.4	3.2	-1.5	3.0	3.2	2.2	2.2	1.5
指标	2000	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
第一产业占 GDP 比重 (%)											
巴西	5.5	5.5	5.1	5.2	5.4	5.3	4.9	5.1	5.3	5.6	5.6
俄罗斯		5.0	4.6	4.4	4.4	4.6	3.8	4.4	3.9	4.0	4.2
印度	25.3	21.6	21.0	21.0	20.4	20.3	21.0	21.6	21.1	20.8	19.5
中国	14.7	11.7	10.7	10.4	10.3	9.9	9.6	9.5	9.5	9.4	9.2
南非	10.7	10.0	10.7	11.3	12.4	11.8	11.9	12.1	11.6	11.3	10.9
指标	2000	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
第二产业占 GDP 比重 (%)											
巴西	26.5	28.6	27.8	27.1	27.4	25.7	27.4	27.2	25.4	24.4	23.4
俄罗斯		38.0	37.2	36.4	35.9	33.7	34.8	37.4	37.0	36.0	35.8
印度	23.7	25.3	26.1	26.3	25.7	25.2	24.3	29.9	28.9	27.9	27.5
中国	45.4	46.9	47.4	46.7	46.8	45.7	46.2	46.1	45.0	43.7	42.6
南非	24.5	23.0	21.3	21.3	22.1	21.6	20.9	20.3	20.5	20.9	21.0
指标	2000	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
第三产业占 GDP 比重 (%)											
巴西	68.0	65.9	67.1	67.7	67.2	69.1	67.8	67.7	69.4	70.0	71.0
俄罗斯		57.0	58.2	59.2	59.7	61.7	61.4	58.2	59.1	60.0	60.0
印度	51.0	53.1	52.9	52.7	53.9	54.5	54.6	48.5	50.0	51.3	53.0
中国	39.8	41.4	41.9	42.9	42.9	44.4	44.2	44.3	45.5	46.9	48.2
南非	64.8	67.0	68.0	67.4	65.5	66.6	67.2	67.5	67.9	67.8	68.0

(来源：国家统计局 2015-07-09)

数据资料来源：

巴西：巴西国家地理与统计局，国民账户体系，2000—2014 年。

俄罗斯：俄罗斯联邦统计局。

印度：1．印度统计和计划执行部中央统计局。2．外币汇率从印度储备银行/印度外汇交易经纪人协会得到。

中国：中华人民共和国国家统计局。

南非：南非统计局和南非储备银行。

统计指数

学习要点

- 指数的意义和分类。
- 总指数的形式与编制方法。
- 运用指数体系进行因素分析。

谈起指数，大家都不陌生（注意：此处的指数不是指数学中的“指数”），在社会经济生活中，各种媒体都有许多关于指数的报道。例如，有反映股票价格变化的股票指数，有通过计算人体身高与体重之间的比值大小来判断是否肥胖的体重指数，有反映企业经营状态的景气指数，以及气象部门的空气污染指数、晨练指数等。与我们的经济和生活关系最密切的是居民消费价格指数（CPI）。但我们经常会感到自己对物价变化的感受与公布的物价指数有反差，实际上，这种反差是我们的心里感觉与科学的统计测量之间的差别。到底什么是指数？它是如何计算的？它表达怎样的含义及如何利用它来进行决策和分析？本章对统计指数的有关理论与方法进行阐述。

8.1 统计指数的概念和种类

8.1.1 统计指数的概念

统计指数（index number）简称指数，它是对社会经济现象进行数量变动分析的一种特殊方法。指数的编制最早起源于物价指数，早在 1650 年英国人沃汉（Rice Vaughan）首创物价指数，用于度量物价的变化情况。其后，指数的应用范围不断扩大，其含义也随之发生了变化。

从指数的含义上看，它有广义与狭义之分。广义指数是指任何两个数值对比形成的相对数。因此，前面章节中我们讨论过的计划完成程度相对数、比较相对数、动态相对数等都可以叫指数。狭义指数是指反映不能直接加总、对比的复杂经济现象总体综合变动的特殊的动态相对数。这种不能直接加总、对比的总体是指数理论着重研究的对象，称为复杂现象总体，反映复杂现象总体综合变动状况的指数称为总指数。本章所讨论的指数是狭义指数。

例如，表 8.1 是某出口公司 2014 年和 2015 年两年三种产品出口量和出口单价资料。根据表中资料，单独看任何一种商品出口单价及出口量，2015 年与 2014 年相比变动的方向和程度

都可以知道（利用普通的动态相对数，即广义的指数）。但我们现在关心的是，从公司的角度来看，出口的三种产品“总的来说”（或“平均来说”）2015 年比 2014 年出口单价及出口量分别是上升还是下降？程度是多少？

表 8.1 某公司商品出口单价和出口量表（1）

商品名称	单位	出口单价（元）		出口量（万）	
		2014 年	2015 年	2014 年	2015 年
跑鞋	双	40	45	120	150
服装	套	65	58	109	140
旅行包	个	180	200	28	22
合 计	—	—	—	—	—

实际上，我们需要求出的是一种特殊的 2015 年与 2014 年比较的动态相对数，它能够反映三种不能直接加总的商品（跑鞋、服装、旅行包）的综合变动，这个特殊的动态相对数就是我们所说的狭义指数。

可见，统计指数具有如下特点：一是相对性。指数是一个相对数，它可以度量总体各变量在不同时间的相对变化。二是综合性。指数能够反映复杂现象的综合变动，这是就狭义指数而言的，也是指数理论和方法的核心问题。没有综合性，指数就不可能发展成为一种独立的理论和方法论体系。三是平均性。统计指数所反映的复杂现象的综合变动，实际上也是各简单现象变动的代表性数值，反映变动的一般水平，因而具有平均的含义。

正是因为统计指数所具有的特点，它在社会经济生活中起到了重要的作用。首先，统计指数能反映复杂现象总体在数量上的变动方向和变动程度。统计指数一般是用百分比表示的相对指标，其比值大于或小于 100%，表示上升或下降的变动方向；具体比 100%大或小的数值，则说明升降变动的相对程度。

其次，统计指数能对现象的总体变动进行因素分析，包括现象总体总量指标、相对指标和平均指标受各个因素变动的影响程度。运用指数分析法，不仅可以从相对数方面分析各构成因素变动对现象总变动的影响程度，还可以从绝对数方面分析其影响的绝对额。

另外，利用连续编制的指数数列，还可以对复杂现象总体长时间发展变化趋势进行分析。这种方法尤其适用于对比分析那些性质不同却有联系的时间数列的变动关系，因为指数数列的对比能够克服不同性质现象不可比的问题。例如，编制零售物价指数数列，来反映一定时期内物价的变动情况；编制每一天的股价指数，来反映股票价格的升与跌，等等。

8.1.2 统计指数的种类

1. 个体指数和总指数

按指数所反映的对象范围的不同，统计指数可分为个体指数、组（或类）指数和总指数。个体指数是反映单个经济现象变动情况的相对数，属于广义指数，如个别产品的物量指数、个

别商品的价格指数等。显然，个体指数是在简单现象总体的条件下存在的。常用的个体指数有：

$$\text{商品销售量个体指数 } k_q = \frac{q_1}{q_0};$$

$$\text{价格个体指数 } k_p = \frac{p_1}{p_0};$$

$$\text{成本个体指数 } k_z = \frac{z_1}{z_0}。$$

式中： q_1 ——报告期销售量；

q_0 ——基期销售量；

p_1 ——报告期价格；

p_0 ——基期价格；

z_1 ——报告期成本；

z_0 ——基期成本；

k ——个体指数。

总指数是反映复杂现象总体在数量上平均变动程度的相对数，用 \bar{k} 表示。例如，综合反映多种产品产量平均变动程度的产品产量总指数 \bar{k}_q ，综合反映多种产品成本平均变动程度的成本总指数 \bar{k}_z ，等等。再如，工业产品总产值指数和零售商品物价总指数等总指数都是在复杂现象总体的条件下进行编制的。它的计算形式有综合指数和平均指数。指数法的应用，要与科学分组相结合，因而在编制总指数的同时，往往还要编制组指数或类指数，借以反映个体内部各部分现象数量上的变动程度。组（或类）指数是反映某一组（或类）指标变动情况的相对数，是相对于个体指数而言的，它实质上还是总指数。

2. 数量指标指数和质量指标指数

按指数所表明指标性质的不同，统计指数可分为数量指标指数和质量指标指数。数量指标指数（index number of quantity）反映研究对象总体总规模的变化程度，如工业产品产量指数、商品销售量指数等；质量指标指数（index number of quality）反映工作质量或成绩好坏、管理水平高低等变动情况，如产品价格指数、劳动生产率指数、产品成本指数等。在指数的应用中，必须重视数量指标指数与质量指标指数的区分，采用不同的编制方法，进行不同情况的动态分析。

3. 定基指数和环比指数

按指数采用基期的不同，统计指数可分为定基指数和环比指数。定基指数指各个时期指数都是采用同一固定时期为基期计算的；环比指数是依次以前一时期为基期计算的。定基指数的基期不依分析时期的变动而变化，可用来反映现象在一个较长时期的变动情况；而环比指数的基期随报告期的变化而变化，可用来反映被研究现象逐期变动的情况。

4. 综合指数和平均数指数

按指数的编制方法不同，统计指数可分为综合指数和平均数指数。综合指数是由两个总量指标对比而形成的指数；平均数指数是以个体指数为基础，采用加权平均方法计算而成的指数。这两种指数形式既各具独立意义，又相互有联系。

5. 动态指数和静态指数

按指数所反映现象的时间不同，统计指数可分为动态指数和静态指数。动态指数是由两个不同时期的同一现象指标值对比而形成的指数；静态指数是指由同一时期不同空间同类现象指标值对比而形成的指数，以及同一空间范围内计划指标与实际指标对比而形成的指数。

6. 现象总体指数和影响因素指数

按指数在指数体系中所处的位置与作用不同，可分为现象总体指数和影响因素指数。现象总体指数是包括两个或两个以上因素同时变动的相对数，属于广义指数。例如，销售额指数，既有销售量的变动，同时也有价格的变动。影响因素指数是只有一个因素变动，并从属于某一现象总体指数的相对数，属于狭义指数。例如，销售量指数，只有销售量一个因素变动，并从属于销售额指数；价格指数，只有价格一个因素变动，并从属于销售额指数。现象总体指数与影响因素指数的关系不能随意形成，而是由现象的客观联系决定的。例如，“销售额=销售量×价格”，由此形成了“销售额指数=销售量指数×价格指数”的关系。这是后述的因素分析的重要理论基础。

8.2 综合指数

总指数的编制有综合指数和平均指数两类方法，综合指数是总指数计算的基本形式和出发点。下面以表 8.1 所列的资料来说明综合指数的编制要点，如表 8.2 所示。

表 8.2 某公司商品出口单价和出口量表（2）

商品名称	计量单位	出口单价（元）		出口单价变动方向	出口量（万）		销售量变动方向
		2014 年	2015 年		2014 年	2015 年	
跑鞋	双	40	45	↑ 上升	120	150	↑ 上升
服装	套	65	58	↓ 下降	109	140	↑ 上升
旅行包	个	180	200	↑ 上升	28	22	↓ 下降
合 计	—	—	—	—	—	—	—

这是一个简化和抽象了的仅由 3 种商品（实际中可能会有成千上万种商品）构成的复杂总体。按照指数的概念，可以得到商品出口量和出口单价两个方面的变动情况。例如，跑鞋，可以根据资料计算得其个体出口量指数为 $k_q=150/120=125\%$ ；个体出口价格指数为 $k_p=45/40=112.5\%$ 。而对反映三种商品出口量和出口价格总的变动情况的相对数即总指数，出于商品的不同属性，使得我们既不能直接将商品的出口量或单价加总起来对比计算总指数，也不能取其个体指数的简单平均数计算总指数。在这里，两个时期的各种商品出口量有增有减，出

口单价有涨有跌，但所研究的这两个特征的变动在总量上又无法直接加总对比得出，这种不能直接加总对比的复杂现象，称为不同度量现象。综合指数的首要问题就是要使不能直接加总的各种商品的数量特征改变为能够加总，进而可以对比的现象总量。商品是使用价值与价值的统一体，商品出口数量总是和出口价格联系在一起的。若将二者结合起来，就能把各种商品的不同使用价值形态变成相同形态的价值量，使得不同度量的总体量变为可以加总，即可以同度量的总体量。

由此引出综合指数编制的两个要点。第一个要点是确定同度量因素，使复杂总体中不能直接加总的量过渡到能直接加总的量。所谓同度量因素，是指在总指数计算的过程中，为解决总体的构成单位及其数量特征不能加总（即不能同度量）的问题而使用的一个媒介因素或转化因素。它有两个作用：一是同度量作用；二是权数作用。第二个要点是需要将相应的同度量因素固定在某一个水平上。例如，为了反映各种商品销售数量总的变动情况，就需要将同度量因素——销售价格固定在某一水平（如基期价格、报告期价格或固定价格水平）上，以便进行对比。如果要反映各种商品销售价格总的变动情况，也需将对应的商品销售量固定在某一水平上。

由于研究社会经济现象有数量指标和质量指标之分，因此综合指数也就有数量指标指数和质量指标指数之别。这两种综合指数编制的基本原理相同，但在编制方法上略有差异，故分别阐述。

8.2.1 数量指标综合指数

现以商品出口量指数为例来说明数量指标综合指数的编制方法和过程。下面利用表 8.1 所列的资料进行分析，思路包括三个方面。

第一，三种商品的计量单位不同，其商品出口量不能直接相加。

第二，利用出口价格作为同度量因素，使不能相加的出口量指标过渡到能够相加的出口额指标。这样就可以将不同商品在两个时期的出口额分别加总，再将两个时期的出口总额进行对比，于是可得出出口额总指数公式：

$$\bar{k}_{qp} = \frac{\sum q_1 p_1}{\sum q_0 p_0} \quad (8-1)$$

第三，为说明出口量的总变动，同度量因素必须使用同一时期的，即假定不同时期的商品出口额是按照同一个时期的出口价格来计算的。

根据总指数的概念和前面三点分析的思路，可以知道数量指标综合指数的一般公式为（下角标 0、1 分别代表基期和报告期）：

$$\bar{k}_q = \frac{\sum q_1 p}{\sum q_0 p} \quad (8-2)$$

在式（8-2）中，综合指数是两个时期现象总量之比。这里是两期出口总额之比，不过其中

只有商品出口量发生变化产生影响，故相比的结果可以说明出口量（即指数化因素）的总变动。这种计算方法总指数采用了一种必要的假定，即假定两个时期的价格相同（不变）来测定出口量的变动。落实使用这个一般公式，同度量因素就必须确定在某一时间上。而商品价格有基期价格、报告期价格、不变价格，将商品价格这个同度量因素固定在其中任何一种，出口量总指数都是成立的，现分别用不同价格作为同度量因素予以分析。

（1）以基期价格作为同度量因素，其出口量总指数的公式为：

$$\bar{k}_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \tag{8-3}$$

式（8-3）称为基期加权综合指数公式，由德国学者拉斯贝尔（Laspeyres）1864 年首次提出，因而也称拉氏公式。

根据表 8.1 所列的资料可以得到所需的对比指标值，如表 8.3 所示。

利用式（8-3）计算商品出口量指数，即

$$\bar{k}_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{19\,060}{16\,925} = 112.61\%$$
$$\sum q_1 p_0 - \sum q_0 p_0 = 19\,060 - 16\,925 = 2\,135 \text{（万元）}$$

表 8.3 商品出口量指数计算表（1）

商品名称	计量单位	出口单价（元） <i>p</i>		出口量（万） <i>q</i>		出口额 <i>pq</i> （万元）			
		2014 <i>p</i> ₀	2015 <i>p</i> ₁	2014 <i>q</i> ₀	2015 <i>q</i> ₁	2014 年实际 <i>p</i> ₀ <i>q</i> ₀	2015 年实际 <i>p</i> ₁ <i>q</i> ₁	假定出口额 <i>p</i> ₁ <i>q</i> ₀	假定出口额 <i>p</i> ₀ <i>q</i> ₁
跑鞋	双	40	45	120	150	4 800	6 750	5 400	6 000
服装	套	65	58	109	140	7 085	8 120	6 322	9 100
旅行包	个	180	200	28	22	5 040	4 400	5 600	3 960
合计	—	—	—	—	—	16 925	19 270	17 322	19 060

式（8-3）的计算结果说明：

- ① 出口量综合变动的方向和程度，即 2015 年出口量比 2014 年增加了 12.61%。
- ② 出口量变动对出口额的影响程度，即变动使 2015 年商品出口额比 2014 年增加了 12.61%。
- ③ 出口量变动对出口额的绝对影响量，即

$$\sum q_1 p_0 - \sum q_0 p_0 = 19\,060 - 16\,925 = 2\,135 \text{（万元）}$$

对公司而言，是多出口了商品使出口额增加了 2 135 万元。

（2）将同度量因素固定在报告期上，计算公式为：

$$\bar{k}_q = \frac{\sum q_1 p_1}{\sum q_0 p_1} \tag{8-4}$$

式(8-4)称为报告期加权综合指数公式,由德国学者派许(Passche)1874年首先提出使用,因而也称为派氏公式。用式(8-4)算得的商品出口量指数为:

$$\bar{k}_q = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{19\,270}{17\,322} = 111.25\%$$

$$\sum q_1 p_1 - \sum q_0 p_1 = 19\,270 - 17\,322 = 1\,948 \text{ (万元)}$$

式(8-4)的计算结果说明:

- ① 出口量综合变动的方向和程度,即总的说来,2015年出口量比2014年增加11.25%。
- ② 出口量变动对出口额的影响程度,即变动使2015年商品出口额比2014年增加了11.25%。
- ③ 出口量变动对出口额的绝对影响量,即

$$\sum q_1 p_1 - \sum q_0 p_1 = 19\,270 - 17\,322 = 1\,948 \text{ (万元)}$$

对公司而言,是多出口了商品使出口额增加了1 948万元。

两个式子所得结果存在的差别是,式(8-3)表示假定价格不变时出口量总的变动情况,式(8-4)表示假定价格已经变化为2015年水平时出口量总的变动情况,后者比前者多了一个影响因素。

(3) 将同度量因素固定在某一固定基期上(即以不变价格作为同度量因素),公式为:

$$\bar{k}_q = \frac{\sum q_1 p_n}{\sum q_0 p_n} \quad (8-5)$$

式中: p_n ——某一特定年份 n 的价格水平。

式(8-5)称为固定加权综合指数公式,由扬格(Young)1818年提出,因而也称为扬格公式。特点是:同度量因素一经选定,不因比较时期的改变而改变。

上述的计算分析说明一个道理:同一数量指标指数的同度量因素,固定在不同时期会得出不同的结果。那么,数量指标指数的同度量因素究竟固定在什么时期为宜呢?这是编制数量指标指数时应该解决的一个重要问题。

编制出口量总指数的目的,在于综合反映多种商品的出口量变动情况,即从总体来说是增加了还是减少了,增加或减少的幅度有多大,以及由此带来的经济效果如何。

在上述3个公式中,式(8-3)(拉氏公式)是假定价格不变,报告期出口总额的计算不受价格变动的影响,因而对比的结果纯粹反映了出口量的变动方向和程度。可见,用基期价格作为同度量因素计算出口量总指数,符合我们的研究目的。用这一指数公式编制定基指数数列时,由于各指数的分母 $\sum q_0 p_0$ 相同,指数间还可以相互比较,便于说明所研究现象变化的程度及其规律性。

式(8-4)(派氏公式)计算的3种商品出口量指数,是报告期实际出口总额与基期出口量

按报告期价格计算的假定出口总额的对比。观察这个指数可发现两个问题：第一，报告期价格 p_1 是由基期价格 p_0 变化而来，用 p_1 作为同度量因素，就把价格变化的影响带入到指数中，使得出口量增长幅度小了。因此，计算出口量总指数时将同度量因素固定在报告期是不甚合理的。第二，指数公式中的分子是报告期出口量按报告期价格计算的出口额，既有随出口量变化而增减的出口额，又有随价格变化而增减的出口额；分母则是由基期出口量按报告期价格计算的假定出口总额。两者的差额除包含有因出口量变化而引起增减的出口额以外，还有受两个时期价格差额影响而增减的出口额。显然，用这一指数公式来分析因出口量变动所带来的实际经济效果是不合理的。

在应用式（8-5）（扬格公式）计算商品出口量总指数时，需注意随着时间的推移，长期将某一时期质量指标作为同度量因素固定下来，可能会脱离客观实际。因此，每隔一段时期就需要调整一次权数，一般是5年或10年更换一次。

综上所述，编制数量指标指数，应以基期的质量指标作为同度量因素，即采用式（8-2）进行计算。编制数量指标指数数列时，应以某一固定时期的质量指标作为同度量因素，即采用式（8-5）进行计算。

8.2.2 质量指标综合指数

质量指标综合指数是说明质量指标总变动情况的动态比较指标。商品价格综合指数、单位生产成本综合指数、劳动生产率综合指数、农作物单位面积产量综合指数等均属于质量指标综合指数。它们的编制原则和应注意解决的问题与数量指标综合指数基本相同，只是处理方法略有差异。下面以商品出口价格综合指数为例，具体说明质量指标综合指数计算公式的形成或编制方法。

按表 8.1 所列资料，测定跑鞋、服装、旅行包 3 种商品的出口价格的个体指数很简单，只要将各种商品的报告期单价与基期单价对比就可得到，即三种商品出口价格的个体指数分别为：

$$k_{\text{跑鞋}} = \frac{p_1}{p_0} = \frac{45}{40} = 112.5\%$$

$$k_{\text{服装}} = \frac{p_1}{p_0} = \frac{58}{65} = 89.23\%$$

$$k_{\text{旅行包}} = \frac{p_1}{p_0} = \frac{200}{180} = 111.11\%$$

计算结果表明：2015 年与 2014 年相比较，跑鞋的价格上涨了 12.5%，服装的价格下跌了 10.77%，旅行包的价格则上涨了 11.11%。

但是，要测定三种商品价格总的变动情况，就要计算价格总指数。那么如何编制三种商品的价格总指数呢？如前所述，由于各种商品的价格反映不同使用价值的实物量的水平，彼此直接相加汇总和对比毫无实际经济意义。但是，若将出口量乘以单价就变成商品的出口额，从而

可以加总对比了。因此，在计算价格总指数时，商品的出口量可以作为同度量因素，固定不变。根据这一原理建立起来的质量指标综合指数的一般公式为：

$$\bar{k}_p = \frac{\sum p_1 q}{\sum p_0 q} \quad (8-6)$$

但是，应将商品出口量固定在基期、报告期还是某一固定期呢？

仿照前述出口量指数的分析思路和计算过程，也可以推出质量指标综合指数的计算公式。

(1) 将同度量因素（出口量）固定在基期上，则出口价格指数公式为：

$$\bar{k}_p = \frac{\sum p_1 q_0}{\sum p_0 q_0} \quad (8-7)$$

式(8-7)与式(8-3)均属于基期加权综合指数公式，即拉氏(Laspeyres)公式。运用式(8-7)，根据表8.3所列资料算出出口价格总的变动是：

$$\bar{k}_p = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{17\,322}{16\,925} = 102.35\%$$

$$\sum p_1 q_0 - \sum p_0 q_0 = 17\,322 - 16\,925 = 397 \text{ (万元)}$$

计算结果表明，在假定基期出口量不变的情况下，报告期三种商品价格总水平比基期上升了2.35%；由于价格总水平上升，使得出口额增加了397万元。

(2) 将同度量因素固定在报告期中，则出口价格指数公式为：

$$\bar{k}_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} \quad (8-8)$$

式(8-8)与式(8-4)均属于报告期加权综合指数公式，即派氏(Passche)公式。运用式(8-8)，根据表8.3所列资料算出出口价格总的变动是：

$$\bar{k}_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{19\,270}{19\,060} = 101.10\%$$

$$\sum p_1 q_1 - \sum p_0 q_1 = 19\,270 - 19\,060 = 210 \text{ (万元)}$$

计算结果表明，在假定报告期出口量不变的情况下，报告期三种商品价格总水平比基期上升了1.10%；由于价格总水平上升，使得出口额增加了210万元。

(3) 将同度量因素固定在某一特定时期上，则出口价格指数公式为：

$$\bar{k}_p = \frac{\sum p_1 q_n}{\sum p_0 q_n} \quad (8-9)$$

式(8-9)与式(8-5)均属于固定加权综合指数公式，即扬格公式。在质量指标综合指数的编制中，作为同度量因素的数量指标往往依据特定的计划要求或某一数量标准来确定。

上述的计算分析说明了这样一个道理：同一质量指标指数的同度量因素，固定在不同时期会得出不同的结果。那么，质量指标指数的同度量因素究竟固定在什么时期为宜呢？这是编制质量指标指数时应该解决的又一重要问题。

实际统计工作中，一般是将报告期的出口量作为同度量因素。这主要是因为以报告期商品出口量作为同度量因素，才能正确反映当前全部商品价格的总变动，使物价指数具有现实的经济意义。如果用基期商品出口量作为同度量因素，就会脱离现实经济生活。

式(8-9)在实际工作中主要用于质量指标的计划完成指数。例如，检查成本计划执行情况时，需要编制成本计划完成指数，其同度量因素是计划数量指标。主要目的在于维护企业计划的严肃性，避免实际情况脱离计划要求。

综上所述，编制价格指数，应将作为同度量因素的出口量固定在报告期。一般认为，编制质量指标指数，应以报告期的数量指标作为同度量因素；编制质量指标的计划完成指数，应以计划数量指标作为同度量因素。

8.3 平均指数

编制综合指数，既可以说明现象变动的方向和程度，又可以说明现象变动所产生的实际效果，其公式计算也比较简单。但编制时却需要具有全面的统计资料，以编制商品价格指数为例，

在应用公式 $\bar{k}_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$ 时，要有各种商品基期、报告期的价格和报告期销售量的对应资料；在

应用公式 $\bar{k}_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}$ 时，要有各种商品基期、报告期的销售量和基期的价格的对应资料。因此，

在某些原始资料不完备的情况下，就不能直接应用综合指数公式，而需要寻求另外的方法计算总指数。平均指数是用非全面资料计算总指数的好方法。

平均指数也称平均数指数，它是从个体指数出发，先计算质量指标和数量指标的个体指数，然后采用加权平均的方法编制总指数。平均数指数与综合指数既有区别，又有联系。两者的联系在于，在一定的权数下，平均数指数是综合指数的一种变形。但是，作为一种独立的指数形式，平均数指数在实际中不仅作为综合指数的变形使用，而且它本身也具有独特的应用价值。

平均指数的计算形式基本上分两种：一种是加权算术平均指数，另一种是加权调和平均指数。

8.3.1 加权算术平均指数

加权算术平均指数是对个体指数按加权算术平均法加以计算，即以个体指数为变量值，以一定时期的总值资料为权数，加权算术平均以计算总指数的方法。

根据指数的定义，个体数量指数和个体质量指数可以分别表示为：

$$k_q = \frac{q_1}{q_0}, \quad k_p = \frac{p_1}{p_0}$$

按照个体指数公式，显然有：

$$q_1=k_qq_0, \quad p_1=k_pp_0$$

由综合指数公式的编制原理与方法可知，总指数计算可按下列公式进行。

（1）使用基期同度量因素的数量指标指数，其公式为：

$$\bar{k}_q = \frac{\sum k_q q_0 p_0}{\sum q_0 p_0} \tag{8-10}$$

（2）使用基期同度量因素的质量指标指数，其公式为：

$$\bar{k}_p = \frac{\sum k_p q_0 p_0}{\sum q_0 p_0} \tag{8-11}$$

就前面的例子来说，式（8-10）和式（8-11）中的 k_q 和 k_p 分别是单一商品的个体出口量和个体价格指数， q_0p_0 则是与个体指数相对应的基期产品出口额。而两个公式的形式，类同加权算术平均数，即 k 是所要平均的变量， qp 是权数。因此，将这种计算总指数的方法称为加权算术平均数指数。

下面利用式（8-10）对表 8.1 中所列的三种商品的出口量变动情况进行分析，如表 8.4 所示。

表 8.4 商品出口量指数计算表（2）

商品 名称	计量 单位	出口单价（元） p		出口量（万） q		个体出口量 指数 k_q （%）	出口额 pq （万元）		个体指数 k_q 与基 期出口额的乘积
		2014 p_0	2015 p_1	2014 q_0	2015 q_1		2014 年实际 p_0q_0	2015 年实际 p_1q_1	
跑鞋	双	40	45	120	150	125	4 800	6 750	6 000
服装	套	65	58	109	140	128.44	7 085	8 120	9 100
旅行包	个	180	200	28	22	78.57	5 040	4 400	3 960
合计	—	—	—	—	—	—	16 925	19 270	19 060

$$\bar{k}_q = \frac{\sum k_q q_0 p_0}{\sum q_0 p_0} = \frac{19\,060}{16\,925} = 112.61\%$$

$$\sum k_q q_0 p_0 - \sum q_0 p_0 = 19\,060 - 16\,925 = 2\,135 \text{（万元）}$$

计算结果表明，三种商品出口量报告期比基期平均增长了 12.61%，由于出口量增长而增加的出口额为 2 135 万元。

因为式（8-10）是按综合指数原理设计的，它在同一资料下与式（8-3）计算的结果完全一致。还需要注意两个问题：

- （1）式（8-10）和式（8-11）均可使用现成的基期总值资料，以方便计算。
- （2）现实中平均数指数计算使用的多为非全面的原始资料，即总体中一部分单位的 k 及对

应的 qp 。若是这样，由于资料范围的不同，它与综合指数计算的结果就会有出入。但如果代表性的个体（商品、产品等）选择得好，则两者不会相差太大，仍可反映总的变动。

8.3.2 加权调和平均指数

如果仅从数学表达式上看，则任何综合指数都可以表达成加权算术平均数的形式。例如，可将式（8-8）写成：

$$\bar{k}_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{\sum k_p p_0 q_1}{\sum p_0 q_1} \tag{8-12}$$

显然，这种变换毫无意义，因为它仍未解决综合指数的资料困难，因此需要变换一种形式来分析。

由 $k_q = \frac{q_1}{q_0}$ 知 $q_0 = \frac{1}{k_q} q_1$ ，由 $k_p = \frac{p_1}{p_0}$ 知 $p_0 = \frac{1}{k_p} p_1$ ，因此有如下计算公式。

（1）使用报告期同度量因素 p 的数量指数，其计算公式为：

$$\bar{k}_q = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{\sum q_1 p_1}{\sum \frac{1}{k_q} q_1 p_1} \tag{8-13}$$

（2）使用报告期同度量因素 q 的质量指数，其计算公式为：

$$\bar{k}_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{\sum p_1 q_1}{\sum \frac{1}{k_p} p_1 q_1} \tag{8-14}$$

式（8-13）和式（8-14）与加权调和平均数的一般形式类似，故将这种总指数的计算方法称为加权调和平均数指数。这里，使用式（8-14）对表 8.1 中所列三种产品的出口价格变动进行计算，如表 8.5 所示。

表 8.5 商品出口价格指数计算表

商品 名称	计量 单位	出口单价（元） p		出口量（万） q		个体出口价格 指数 k_q （%）	出口额 pq （万元）		报告期出口额除以 个体指数 $p_1 q_1 / k_p$
		2014 p_0	2015 p_1	2014 q_0	2015 q_1		2014 年实际 $p_0 q_0$	2015 年实际 $p_1 q_1$	
跑鞋	双	40	45	120	150	112.5	4 800	6 750	6 000
服装	套	65	58	109	140	89.23	7 085	8 120	9 100
旅行包	个	180	200	28	22	111.11	5 040	4 400	3 960
合计	—	—	—	—	—	—	16 925	19 270	19 060

$$\bar{k}_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{\sum p_1 q_1}{\sum \frac{1}{k_p} p_1 q_1} = \frac{19\,270}{19\,060} = 101.10\%$$

$$\sum p_1 q_1 - \sum \frac{1}{k_p} p_1 q_1 = 19\,270 - 19\,060 = 210 \text{ (万元)}$$

计算结果表明，三种商品出口价格报告期比基期平均增长了 1.10%，由于出口价格增长而增加的出口额为 210 万元。这个结果与采用综合指数公式计算的结果完全相同。

8.4 指数体系及因素分析

通过指数的编制方法看到，一个指数通常只能说明某一方面的问题，因此在实践中往往需要将多个指数结合起来形成相应的指数体系，以便对相互联系的经济现象作深入分析。

8.4.1 指数体系的含义与作用

1. 指数体系的含义

社会经济现象的变动往往受多种因素的影响，各种因素不是孤立存在的，而是相互联系和相互影响的。例如，反映全国居民生活水平状况，可以用职工工资总额指数、人均可支配收入指数、人均纯收入指数、居民消费水平指数、居民消费价格指数等。这些指数从不同侧面反映居民收入、消费状况，具有一定的内在联系。可以说，这些指数构成了反映居民生活状况的指数体系。又如，工业总产值是由产品产量和产品出厂价格两个因素构成的，商品销售额是由商品销售量和销售价格两个因素构成的，产品利润是由生产量、产品价格和销售利润率构成的，等等。因而，工业总产值、商品销售额、产品利润等的变动，必然会受到各构成因素的影响。于是，我们将那些经济上有联系，在数量上保持一定关系的若干指数形成的整体，称作指数体系。上述例子，可用指数体系来反映它们的联系，即

$$\text{工业总产值指数} = \text{产品产量指数} \times \text{出厂价格指数}$$

$$\text{商品销售额指数} = \text{商品销售量指数} \times \text{物价指数}$$

$$\text{销售利润指数} = \text{销售量指数} \times \text{销售价格指数} \times \text{销售利润率指数}$$

指数体系所反映的是客观事物本身的联系，因而在编制指数体系时，应以综合指数的一般原理为依据，即编制数量指标指数时，以基期质量指标作为同度量因素；编制质量指标指数时，则以报告期的数量指标作为同度量因素。

值得注意的是，组成指数体系的指数必须满足两个条件：各因素指数的乘积应等于总变动指数；各因素指数分子、分母差额总和应等于总量指标指数实际发生的总差额。

2. 指数体系的作用

指数体系的主要作用表现在两个方面。

(1) 进行因素分析。因素分析是利用指数体系从数量方面分析现象总变动中，各影响因素对其影响的方向、程度及绝对效果。为了更好地应用这一方法，在进行因素分析时，应注意以下几个问题。

首先,因素分析是以指数体系为基本依据。在指数体系中,总变动指数与影响因素指数数量关系表现在两方面:一是从相对数上看,总变动指数等于各影响因素指数的乘积;二是从绝对数上看,总变动指数分子与分母的差额等于各影响因素分子与分母差额之和。

其次,当测定某一因素变动影响时,必须将其他因素固定下来,固定的方法以综合指数编制的一般原理为依据,即测定数量指标因素变动影响时,将作为同度量因素的质量指标固定在基期;测定质量指标因素变动影响时,将作为同度量因素的数量指标固定在报告期。需要注意的是,在进行多因素分析时,判断各影响因素指标是数量指标还是质量指标是两两相比较而言的。例如,有指数体系如下:

$$\text{原材料费用总额指数} = \text{产品产量指数} \times \text{单位产品原材料消耗量(单耗)指数} \times \text{单位原材料价格指数(单价)}$$

在这个指数体系中,产品产量与单耗比较,产品产量是数量指标,单耗是质量指标;而单耗与单价比较,单耗又成了数量指标,单价则是质量指标。

最后,在进行多因素分析时,要注意各影响因素指数的合理排序问题。一般是数量指标指数在先,质量指标指数在后。

(2) 进行指数间的推算。根据指数体系中各指数间的相互关系,当已掌握指数体系中若干个指数,对剩下的其中某个未知指数就可运用指数体系的关系进行推算。例如,已知2015年与2014年相比,某地区物价综合上涨了5%,销售额上涨了10%,则销售量指数为 $110\% \div 105\% = 104.76\%$,表明2015年与2014年相比销售量增加了4.76%。

8.4.2 因素分析应用举例

因素分析是依据指数体系的理论,分析多因素影响的社会经济现象总变动中,各因素的影响方向和程度的方法。

进行因素分析一般有四个步骤:

第一步,分析被研究对象及其影响因素。这里的“被研究对象”是具体的统计指标,如商品销售额、流通过费用额、原材料费用总额等。当明确了被研究现象是某个统计指标时,就要分析这个统计指标含有哪些影响因素,这是因素分析的基础。

第二步,建立指数体系。相对数关系式表现为现象总体指数等于各影响因素指数的乘积;绝对数关系式表现为现象总体指数分子与分母的差额等于各影响因素分子与分母差额之和。

第三步,搜集资料,计算指数体系两个关系式中各项数值。按公式内容和要求,收集有关资料,并进行整理、计算。

第四步,根据计算结果,做出分析结论和简要的文字说明。

利用指数体系进行因素分析,依据分析对象不同可分为对总量指标、相对指标和平均指标变动进行因素分析,下面分别说明。

1. 总量指标变动的因素分析

总量指标变动的因素分析可按其影响因素的多少不同，分为两因素分析和多因素分析。

1) 两因素分析

总量指标的两因素分析，有相对数和绝对额分析。

$$\text{相对数分析: } \frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1} \tag{8-15}$$

$$\text{绝对额分析: } \sum q_1 p_1 - \sum q_0 p_0 = (\sum q_1 p_0 - \sum q_0 p_0) + (\sum p_1 q_1 - \sum p_0 q_1) \tag{8-16}$$

【实例 8.1】 以表 8.6 所列资料对销售额的变动进行因素分析。

表 8.6 商品销售额变动的因素分析表

商品名称	计量单位	价格 p (元)		销售量 q		销售额 pq		
		基期 p_0	报告期 p_1	基期 q_0	报告期 q_1	基期实际 $p_0 q_0$	报告期实际 $p_1 q_1$	假定销售额 $p_0 q_1$
甲	米	25	20	400	600	10 000	12 000	15 000
乙	件	40	36	500	600	20 000	21 600	24 000
丙	公斤	50	60	200	180	10 000	10 800	9 000
合计	—	—	—	—	—	40 000	44 400	48 000

第一步：列出分析对象。报告期与基期相比较销售额的变化，即：

$$\bar{k}_{qp} = \frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{44\,400}{40\,000} = 111\%$$

$$\sum q_1 p_1 - \sum q_0 p_0 = 44\,400 - 40\,000 = 4\,400 \text{ (元)}$$

第二步：建立指数体系。

$$\frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

$$\sum q_1 p_1 - \sum q_0 p_0 = (\sum q_1 p_0 - \sum q_0 p_0) + (\sum p_1 q_1 - \sum p_0 q_1)$$

第三步：进行因素分析。

$$\text{销售量总指数: } \bar{k}_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{48\,000}{40\,000} = 120\%$$

$$\sum q_1 p_0 - \sum q_0 p_0 = 48\,000 - 40\,000 = 8\,000 \text{ (元)}$$

所以，由于销售量的变化，使销售额上升了 20%，绝对额增加了 8 000 元。

$$\text{销售价格总指数: } \bar{k}_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{44\,400}{48\,000} = 92.5\%$$

$$\sum p_1q_1 - \sum p_0q_1 = 44\,400 - 48\,000 = -3\,600 \text{ (元)}$$

所以，由于销售价格的变化，使销售额下降了 7.5%，绝对额减少了 3 600 元。

第四步：综合影响。

$$111\% = 120\% \times 92.5\%$$

$$4\,400 = 8\,000 + (-3\,600)$$

计算结果表明，销售额增长 11%是销售量上升 20%与价格下降 7.5%共同作用的结果；销售额增加 4 400 元，是销售量上升使销售额增加 8 000 元、价格下降使销售额减少 3 600 元这两种因素共同作用的结果。

【实例 8.2】 某商业企业有关资料如表 8.7 所示。

表 8.7 某商业企业工资及职工人数资料

指 标	2014 年	2015 年
工资总额（万元）	3 600	4 875
职工人数（人）	500	625
平均工资（万元/人年）	7.2	7.8

分析该企业工资总额的变动受职工人数和平均工资变动的影响情况，用 x 表示平均工资， f 表示职工人数，则

$$\text{工资总额指数} = \frac{\bar{x}_1 f_1}{\bar{x}_0 f_0} = 4\,875 / 3\,600 = 135.42\%$$

$$\text{工资总额增加的绝对数额} = \bar{x}_1 f_1 - \bar{x}_0 f_0 = 4\,875 - 3\,600 = 1\,275 \text{ (万元)}$$

$$\text{职工人数指数} = \frac{f_1}{f_0} = \frac{625}{500} = 125\%$$

$$\text{职工人数增长使工资总额增加的绝对额} = (f_1 - f_0) \bar{x}_0 = (625 - 500) \times 7.2 = 900 \text{ (万元)}$$

$$\text{平均工资指数} = \frac{\bar{x}_1}{\bar{x}_0} = 7.8 / 7.2 = 108.33\%$$

$$\text{职工人数增长使工资总额增加的绝对额} = (\bar{x}_1 - \bar{x}_0) f_1 = (7.8 - 7.2) \times 625 = 375 \text{ (万元)}$$

三个指数之间的关系为：

$$135.42\% = 125\% \times 108.33\%$$

$$1\,275 = 900 + 375$$

计算结果表明，2015 年与 2014 年相比较，该企业工资总额增长了 35.42%，增加的绝对数额为 1 275 万元。其中，职工人数增长了 25%，由于职工人数增长使工资总额的绝对数增加了 900 万元；平均工资提高了 8.33%，由于平均工资提高，使工资总额的绝对数额增加了 375 万元。

2) 多因素分析

以上我们运用指数体系对总量指标的变动进行了两因素分析，这种分析方法还可以推广运用于3个或3个以上的因素分析。例如，工业产品原材料支出总额变动可分解为产量、单位产品原材料消耗量和单位原材料价格3个因素的变动影响。因此，就需要编制原材料支出额指数及其所包括的3个因素指数的指数体系，来进行多因素变动的联系分析。又如，工业总产值动态变动可以分解为对职工人数、工人数占职工人数比重和工人劳动生产率3个因素进行变动影响分析，利税可以分解为销售量、销售价和利税率3个因素，等等。多因素的分析方法和两因素的分析方法基本原理是相同的，但由于多因素现象指数体系所包括的现象因素较多，指数的编制过程比较复杂。

因此，在进行多因素现象分析时，必须注意以下两点：

① 在编制多因素指标所组成的综合指数时，为了测定某一因素指标的变动影响，要把两个或两个以上因素固定不变。这里仍然利用综合指数编制的一般原理，来确定固定因素所属时期，即在反映数量指标因素的变动影响时，应以基期的质量指标为同度量因素；而在反映质量指标因素的变动影响时，应以报告期数量指标为同度量因素。同时，由于所包括的因素较多，还要考虑多因素的合理排列顺序，来确定同度量因素。

② 对综合指数中的多因素排列顺序，要具体分析现象总体的内容，依据现象因素的联系加以具体确定。现以工业产品原材料支出额的组成因素排列顺序为例来加以说明，依据它们之间的联系，要按产量、单位产品原材料消耗量、单位原材料价格的顺序排列，即：

$$\begin{array}{c} \text{原材料消耗量} \\ \text{原材料支出总额} = \text{产品产量} \times \text{原材料单耗} \times \text{原材料单价} \\ \text{单位产品原材料消耗额} \end{array}$$

这样，产品产量相对于单位产品原材料消耗额是数量指标，原材料单价相对于原材料消耗用量来说是质量指标，这与前面讲的两因素分析法一样了。这样，就可以编制多个因素指数，用于分析这些因素变动对现象总体变动的影响作用。

下面以工业产品原材料支出额分解为3个因素的资料为例进行说明。设 m 为原材料单位消耗量， q 、 p 表示产量和原材料价格，则工业产品原材料支出额为 qmp ，各因素指数所形成的指数体系（相对数）为：

$$\underbrace{\frac{\sum q_1 m_1 p_1}{\sum q_0 m_0 p_0}}_I = \underbrace{\frac{\sum q_1 m_0 p_0}{\sum q_0 m_0 p_0}}_{II} \times \underbrace{\frac{\sum q_1 m_1 p_0}{\sum q_1 m_0 p_0}}_{III} \times \underbrace{\frac{\sum q_1 m_1 p_1}{\sum q_1 m_1 p_0}}_{IV} \quad (8-17)$$

其中：I 对应于原材料支出总额指数；

II 对应于产量指数；

III 对应于单位产品原材料消耗指数；

IV 对应于单位原材料价格指数。

$$\sum q_1 m_1 p_1 - \sum q_0 m_0 p_0 = (\sum q_1 m_0 p_0 - \sum q_0 m_0 p_0) + (\sum q_1 m_1 p_0 - \sum q_1 m_0 p_0) + (\sum q_1 m_1 p_1 - \sum q_1 m_1 p_0) \tag{8-18}$$

【实例 8.3】 以表 8.8 中所列资料为例，针对某企业生产三种产品的主要原材料的支出额及其分解因素资料，编制指数体系并进行多因素分析。

表 8.8 产品原材料费用额变动的因素分析表

商品名称	原材料	产量 q		单耗 m		原材料单价（元）		原材料费用额（元）			
		q_0	q_1	m_0	m_1	p_0	p_1	$q_0 m_0 p_0$	$q_1 m_0 p_0$	$q_1 m_1 p_0$	$q_1 m_1 p_1$
甲（个）	公斤	600	900	1.0	0.9	15.0	18.0	9 000	13 500	12 150	14 580
乙（件）	米	1 500	1 800	2.2	1.9	18.0	20.0	59 400	71 280	61 560	68 400
丙（只）	米	900	900	2.5	2.0	25.0	24.0	56 250	56 250	45 000	43 200
合计	—	—	—	—	—	—	—	124 650	141 030	118 710	126 180

解：

（1）原材料费用总额指数：

$$\bar{k}_{qmp} = \frac{\sum q_1 m_1 p_1}{\sum q_0 m_0 p_0} \times 100\% = \frac{126\,180}{124\,650} = 101.23\%$$
$$\sum q_1 m_1 p_1 - \sum q_0 m_0 p_0 = 126\,180 - 124\,650 = 1\,530 \text{（元）}$$

计算结果说明，该企业报告期生产的三种产品的 3 种原材料的费用额比基期增长了 1.23%，绝对数额增加了 1 530 元。

（2）分析各因素对总变动的影响情况，分别计算 3 个因素指数。

① 产量总指数：

$$\bar{k}_q = \frac{\sum q_1 m_0 p_0}{\sum q_0 m_0 p_0} = \frac{141\,030}{124\,650} = 113.14\%$$
$$\sum q_1 m_0 p_0 - \sum q_0 m_0 p_0 = 141\,030 - 124\,650 = 16\,380 \text{（元）}$$

计算结果说明，由于报告期产量比基期增加，使原材料总费用额上升了 13.14%，其绝对额增加了 16 380 元。

② 原材料单位消耗总指数：

$$\bar{k}_m = \frac{\sum q_1 m_1 p_0}{\sum q_1 m_0 p_0} = \frac{118\,710}{141\,030} = 84.17\%$$
$$\sum q_1 m_1 p_0 - \sum q_1 m_0 p_0 = 118\,710 - 141\,030 = -22\,320 \text{（元）}$$

计算结果说明，报告期产品原材料的单位消耗比基期下降了 15.83%，由于单耗减少，使原材料费用额减少了 22 320 元。

③ 原材料价格总指数：

$$\bar{k}_p = \frac{\sum q_1 m_1 p_1}{\sum q_1 m_1 p_0} = \frac{126\,180}{118\,710} = 106.29\%$$

$$\sum q_1 m_1 p_1 - \sum q_1 m_1 p_0 = 126\,180 - 118\,710 = 7\,470 \text{ (元)}$$

计算结果说明，由于报告期原材料价格上涨，使原材料费用支出额增加了 6.29%，增加支出费用为 7 470 元。

以上 4 种指数关系，满足 $\bar{k}_{qmp} = \bar{k}_q \times \bar{k}_m \times \bar{k}_p$

$$\sum q_1 m_1 p_1 - \sum q_0 m_0 p_0 = (\sum q_1 m_0 p_0 - \sum q_0 m_0 p_0) + (\sum q_1 m_1 p_0 - \sum q_1 m_0 p_0) + (\sum q_1 m_1 p_1 - \sum q_1 m_1 p_0)$$

亦即：101.23%=113.14%×84.17%×106.29%

$$1\,530 = 16\,380 + (-22\,320) + 7\,470$$

2. 相对指标变动的因素分析

相对数是由两个指标对比得到的，这两个指标的变动都会影响相对数的变动，故亦可进行因素分析。因为 $C=A/B$ （设 A 为数量指标， B 为质量指标），所以

$$\frac{C_1}{C_0} = \frac{A_1/B_1}{A_0/B_0} = \frac{A_1/B_0}{A_0/B_0} \times \frac{A_1/B_1}{A_1/B_0} \quad (8-19)$$

$$A_1/B_1 - A_0/B_0 = (A_1/B_0 - A_0/B_0) + (A_1/B_1 - A_1/B_0) \quad (8-20)$$

【实例 8.4】 以表 8.9 中所列资料为例，针对某公司商品销售和库存资料，编制相对指标指数体系并进行因素分析。

表 8.9 某公司商品销售和库存资料

指 标	符 号	2016 年第 1 季度	2015 年第 1 季度	指数%
商品销售额（万元）	A	A ₁ 3 609.00	A ₀ 2 755.00	131.00
平均库存额（万元）	B	B ₁ 1 432.00	B ₀ 1 450.00	98.76
商品周转次数（次）	C	C ₁ 2.52	C ₀ 1.90	132.63

根据表中资料计算：A₁/B₁=2.52（次）；A₀/B₀=1.90（次）；

$$A_1/B_0 = 3\,609/1\,450 = 2.49 \text{ (次)}$$

所以

$$\frac{2.52}{1.90} = \frac{2.49}{1.90} \times \frac{2.52}{2.49}$$

$$132.63\% = 131.05\% \times 101.20\%$$

$$2.52 - 1.90 = (2.49 - 1.90) + (2.52 - 2.49)$$

$$0.62 \text{ 次} = 0.59 \text{ 次} + 0.03 \text{ 次}$$

计算结果表明, 商品周转次数 2016 年第 1 季度与上年同季度比较增长了 32.63%, 这是因为销售额变动使之增长 31.05%, 平均库存额变动使之增长 1.20%; 商品周转次数增加了 0.62 次, 是因销售额变动增加 0.59 次和因平均库存额变动增加 0.03 次两因素共同作用的结果。

3. 平均指标变动的因素分析

统计指数法还可用于平均指标的动态分析。平均指标指数就是用来反映总平均数变动及其原因的相对数, 即同一经济内容两个不同时期的平均指标数值之比。它是对平均指标变动情况进行分析的一种方法。平均指标指数的一般公式为:

$$k = \frac{\bar{x}_1}{\bar{x}_0} \quad (8-21)$$

式中: \bar{x}_1 ——报告期平均指标数值;

\bar{x}_0 ——基期平均指标数值。

常见的平均指标指数有: 劳动生产率指数、平均工资指数、平均单位成本指数等。下面以劳动生产率指数为例, 进一步介绍平均指标指数的含义及基本原理。

劳动生产率指数的公式为:

$$k = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\frac{\sum x_1 f_1}{\sum f_1}}{\frac{\sum x_0 f_0}{\sum f_0}} = \frac{\sum x_1 \square \frac{f_1}{\sum f_1}}{\sum x_0 \square \frac{f_0}{\sum f_0}} \quad (8-22)$$

式中: $\sum f_0$ 和 $\sum f_1$ 分别代表基期和报告期的工人总人数; $\sum x_0 f_0$ 和 $\sum x_1 f_1$ 分别代表基期和报告期的产量; x_0 和 x_1 代表基期和报告期的劳动生产率, \bar{x}_0 和 \bar{x}_1 分别代表基期和报告期的平均劳动生产率; f 或 $\frac{f}{\sum f}$ 是一种结构相对数, 简称结构。由此可见, 劳动生产率指数反映两个因素变动的影响, 即各组工人劳动生产率变动的影响和各组工人人数在全部工人总数中所占比重变动的影响。这是因为加权算术平均数的大小取决于变量标志值和权数。因此, 平均指标指数所反映的变动程度, 也包括两个因素间的影响。为了测定一个因素的变动情况, 必须将另一个因素固定下来。

平均指标指数分析需要编制 3 种平均指标指数, 它们是可变构成指数、固定构成指数和结构变动影响指数, 并形成如下的指数体系:

$$\text{可变构成指数} = \text{固定构成指数} \times \text{结构变动影响指数}$$

它们分别简称为可变指数、固定指数和结构指数。下面分别进行介绍。

(1) 可变指数。可变指数是报告期平均指标与基期平均指标之比。例如, 在进行总平均工资变动的因素分析时, 首先要计算总平均工资指数, 计算总平均工资指数的报告期和基期总平均工资, 是分别以各个时期工人数为权数对各组工资水平进行平均计算的。所以, 两期总平均工资的变动, 不仅反映了各组(类)工资水平的变动, 而且受各组工人数结构变动的影响。这种包括结构变动影响作用的总平均工资指数, 就是平均工资可变构成指数, 其公式如下:

$$\text{平均工资可变构成指数 } k_{\text{可变}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 f_1}{\sum f_1} \div \frac{\sum x_0 f_0}{\sum f_0} \quad (8-23)$$

(2) 固定指数。固定指数是假定构成不变(即 $\frac{f}{\sum f}$ 频率不变), 纯粹反映组平均数总变动的相对数。例如, 在分析企业总平均工资变动中各组工资水平变动影响关系时, 依据综合指数编制的原理, 为了消除结构因素的变动影响、反映各组工资水平的变动程度, 要把工人数加以固定, 并且固定在报告期上。这种工人数结构固定的总平均工资指数, 称为平均工资的固定指数, 其公式如下:

$$\text{平均工资固定指数 } k_{\text{固定}} = \frac{\sum x_1 f_1}{\sum f_1} \div \frac{\sum x_0 f_1}{\sum f_1} \quad (8-24)$$

这一指数的经济内容是: 分子指标是报告期企业实际平均工资; 分母指标是假定各组工资水平保持不变的情况下, 企业报告期平均工资。它们之间的差别, 只是由两期各组工人工资水平变动所引起的。

(3) 结构指数。结构指数是指假定组平均数不变(即 x_0 不变), 纯粹反映结构变动程度的相对数。例如, 分析工人变动影响指数, 必须将各组工人工资水平因素固定起来, 并将它固定在基期水平上。其计算公式如下:

$$\text{平均工资结构变动影响指数 } k_{\text{结构}} = \frac{\sum x_0 f_1}{\sum f_1} \div \frac{\sum x_0 f_0}{\sum f_0} \quad (8-25)$$

这个指数的分子指标是假定各组工人工资水平保持不变的情况下, 企业报告期的平均工资; 分母指标是基期企业实际平均工资。它们的对比关系, 可以表明两期中各级工人数构成变动对企业平均工资变动的影响。

(4) 平均指标指数体系。通过以上分析我们看到, 平均指标变动的因素分析, 实质上是现象结构的变动分析, 因为分析所使用的指数都与结构有关。也就是说, 可变指数是包含了结构变动因素的平均指标指数, 固定指数是排除了结构变动影响的平均指标指数, 而结构指数是纯粹考虑了结构变动影响的平均指标指数。上述各种指数之间具有一定的内在联系, 形成一个指数体系, 即

相对数为

$$\frac{\sum x_1 f_1}{\sum f_1} \div \frac{\sum x_0 f_0}{\sum f_0} = \left(\frac{\sum x_0 f_1}{\sum f_1} \div \frac{\sum x_0 f_0}{\sum f_0} \right) \times \left(\frac{\sum x_1 f_1}{\sum f_1} \div \frac{\sum x_0 f_1}{\sum f_1} \right) \quad (8-26)$$

绝对数为

$$\frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0} = \left(\frac{\sum x_0 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0} \right) + \left(\frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_1}{\sum f_1} \right) \tag{8-27}$$

通过平均指标指数体系，用具体例子来说明平均指标指数的分析方法及其应用。

【实例 8.5】 根据表 8.10 中的某企业工人人数、工资额资料，说明平均指标指数的分析及应用。

表 8.10 某企业职工人数和工资水平资料

职工类别	工人数（人）		平均工资（元/人月）		工资总额（元）		
	基期 f_0	报告期 f_1	基期 x_0	报告期 x_1	基期 $\sum x_0 f_0$	报告期 $\sum x_1 f_1$	假定 $\sum x_0 f_1$
老职工	520	320	4 860	5 100	2 527 200	1 632 000	1 555 200
新职工	470	720	2 560	2 680	1 203 200	1 929 600	1 843 200
合计	990	1 040	—	—	3 730 400	3 561 600	3 398 400

从资料中可以看出，该企业新老职工的平均工资报告期比基期都提高了，其中老职工人均提高 240 元，新职工人均提高 120 元。但全厂职工的平均工资是否也提高呢？下面通过计算来分析说明该企业总平均工资的变动情况及原因。

根据表中资料求得：

$$\begin{aligned} \bar{x}_0 &= \frac{\sum x_0 f_0}{\sum f_0} = \frac{3\,730\,400}{990} = 3\,768.08 \text{（元）} \\ \bar{x}_1 &= \frac{\sum x_1 f_1}{\sum f_1} = \frac{3\,561\,600}{1\,040} = 3\,424.62 \text{（元）} \\ k_{\text{可变}} &= \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 f_1}{\sum f_1} \div \frac{\sum x_0 f_0}{\sum f_0} = \frac{3\,561\,600}{1\,040} \div \frac{3\,730\,400}{990} = 90.89\% \\ \frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0} &= 3\,424.62 - 3\,768.08 = -343.46 \text{（元）} \end{aligned}$$

这说明该企业工人总平均工资报告期比基期减少了 9.11%，减少的金额为 343.46 元。

$$\begin{aligned} k_{\text{固定}} &= \frac{\sum x_1 f_1}{\sum f_1} \div \frac{\sum x_0 f_1}{\sum f_1} = \frac{3\,561\,600}{1\,040} \div \frac{3\,398\,400}{1\,040} = 104.80\% \\ \frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_1}{\sum f_1} &= 3\,424.62 - 3\,267.69 = 156.93 \text{（元）} \end{aligned}$$

这说明该企业工人组平均工资水平报告期比基期总的上升 4.80%，由此使总平均工资增加 156.93 元。

$$k_{\text{结构}} = \frac{\sum x_0 f_1}{\sum f_1} \div \frac{\sum x_0 f_0}{\sum f_0} = 3\,267.69 \div 3\,768.08 = 86.72\%$$

$$\frac{\sum x_0 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0} = 3\,267.69 - 3\,768.08 = -500.39 \text{ (元)}$$

计算结果表明，由于该企业工人人数结构的变动使总平均工资减少了 13.28%，减少金额平均为每个职工 500.39 元。

现将上面分析计算结果用平均指标指数体系表示出来，为：

$$\frac{\sum x_1 f_1}{\sum f_1} \div \frac{\sum x_0 f_0}{\sum f_0} = \left(\frac{\sum x_0 f_1}{\sum f_1} \div \frac{\sum x_0 f_0}{\sum f_0} \right) \times \left(\frac{\sum x_1 f_1}{\sum f_1} \div \frac{\sum x_0 f_1}{\sum f_1} \right)$$

$$90.89\% = 104.80\% \times 86.72\%$$

$$\frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0} = \left(\frac{\sum x_0 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0} \right) + \left(\frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_1}{\sum f_1} \right)$$

$$-343.46 = 156.93 + (-500.39)$$

综上，该企业全体工人平均工资报告期为基期的 90.89%，比基期减少了 9.11%，这是由于各组工人工资水平上升使总平均工资提高了 4.80%，由于工人人数结构的变化使平均工资减少了 13.28%，两者共同作用使平均工资降低了 9.11%。每个工人的平均工资减少额为 343.46 元，是由于各组工资水平上升使其增加 156.93 元，工人结构变化使其减少了 500.39 元，这两个因素共同作用所致。

8.5 常用价格指数简介

8.5.1 消费者价格指数

消费者价格指数（consumer price index），英文缩写为 CPI，是反映与居民生活有关的产品及劳务价格统计出来的物价变动指标，一般以百分比为表达形式。

国际上公认 CPI 有三大用途，即通货膨胀指标、国民经济核算及支付的调整。

从 2001 年起，我国采用国际通用做法，逐月编制并公布以 2000 年价格水平为基期的居民消费价格定基指数，作为反映我国通货膨胀（或紧缩）程度的主要指标。居民消费价格指数可按城乡分别编制城市居民消费价格指数和农村居民消费价格指数，也可按全社会编制全国居民消费价格总指数。经国务院批准，国家统计局城调总队负责全国居民消费价格指数的编制及相关工作，并组织、指导和管理各省、区、市的消费价格调查统计工作。

我国编制价格指数的商品和服务项目，根据全国城乡近 13 万户居民家庭消费支出构成资料 and 有关规定确定，目前共包括食品、烟酒及用品、衣着、家庭设备用品及服务、医疗保健及个

人用品、交通和通信、娱乐教育文化用品及服务、居住 8 大类，262 个基本分类，约 700 个代表品种。目前参加全国数据汇总的调查市、县约 550 个，调查网点包括超市、菜市场、百货商场、医院、旅行社等约 6.3 万个。

居民消费价格指数就是在对采价点进行价格调查的基础上，根据国际规范的流程和公式算出来的。在美国构成该指标的主要商品共分七大类，其中包括：食品、酒和饮品，住宅，衣着，交通，医药健康，娱乐，其他商品及服务。在美国，消费物价指数由劳工统计局每月公布，有两种不同的消费物价指数。一是工人和职员消费物价指数，简称 CPW；二是城市消费者的消费物价指数，简称 CPIU。

编制居民消费价格指数需要两类资料：一是各项消费商品和服务的价格变动数据；二是各项消费商品和服务在居民生活中的开支权重。为此，我们必须选择好物价调查点和调查代表规格品。

编制 CPI 简单地说，可以概括为三个步骤。

第一，确定基本分类中每个指数商品的代表规格品。目前计算 CPI 的代表规格品既包括商品消费，也包括服务消费；既包括传统消费项目，也包括汽车等“现代消费”。大部分基本分类选择 2~3 个规格品，目前我国编制指数的地区根据其规模选择了 550~750 个代表规格品。这些代表规格品的选择是根据近 4 万户城市居民家庭和 6 万多户农村居民的消费支出构成资料和有关规定决定的。一般而言，大城市的代表规格品多些，小城镇就少些。

规格品的选取并不是随意确定的，确定规格品也有 3 个原则：一是各调查市、县结合当地的实际情况，选择居民消费量大、市场供应相对稳定、价格变动趋势有代表性的规格品；二是选中的规格品之间，性质相隔越远越好，即价格变动特征的相关性愈低愈好；三是选中的工业消费品必须是合格产品，产品包装上有注册商标、产地、规格等级等标识。例如，电视机是规格品，每个地方要选 3~5 个规格，各地根据经济水平的不同选择的规格也不同，上海可能选取的是背投大屏幕，而西北省份则可能选取 25 英寸为代表规格品。

第二，采集代表规格品的价格。价格调查采用定人、定点、定时直接调查的方法。为保证价格资料的真实性、可靠性和准确性，国家统计局要求各选中的调查市、县必须做到，采价时不受挂牌价格的限制，要采集居民在商店或农贸实际成交的价格；对与居民生活密切相关、价格变动比较频繁的商品，至少每 5 天调查一次；一般性商品每月调查 2~3 次；由国家控制价格的一些主要商品、服务项目或价格变动相对稳定的商品、服务项目，可视情况每月或每季调查一次。

第三，进行价格指数的计算。价格指数的计算包括三个步骤：一是代表规格品平均价格的计算；二是基本分类指数的计算，包括月环比指数的计算等；三是类别及总指数逐级加权平均计算。

世界各国的消费价格指数多用加权算术平均法计算，我国也采用加权算术平均法。其公式为 $\bar{k}_p = \sum k_p W$ ，式中 \bar{k}_p 代表总（类）指数， k_p 代表个体（类）指数， W 代表权数。权数是按总指数、大类、中类、小类和代表品分层计算的，每层的权数总和 $\sum W$ 为 100。各省的权数按

中选的样本市、县的资料计算，全国的权数则根据各省的资料计算。

计算时，先计算商品或服务项目的个体指数，然后计算小类、中类、大类指数，最后计算总指数。由小到大，逐步升级，层层平均。

现以城市消费价格指数为例计算如下（见表 8.11）。

表 8.11 某城市居民消费价格指数
(年 月 / 年 月)

城市编码:

商品类别及品名	规格等级牌号	计量单位	平均价格（元）		权数	以上年为基期	
			上年	本年		指数（%）	指数×权数
（甲）	（乙）	（丙）	（1）	（2）	（3）	（4）	（5）=（4）×（3）
总指数					100	114.13	—
一、食品类					《34》	114.41	3 889.94
1. 粮食					<20>	10.401	2 208
（1）细粮					（88）	110.52	9 725.46
面粉	普通粉	千克	4.16	4.74	38	113.94	4 329.72
大米	标二	千克	4.58	4.97	55	108.52	5 968.6
江米	标二	千克	5.20	5.59	5	107.50	537.5
挂面	富强粉	千克	4.56	4.93	2	108.11	216.22
（2）粗粮					（12）	109.53	1 314.36
2. 淀粉及薯类					<1>	111.02	111.02
3. 干豆类及豆制品					<2>	119.06	238.12
4. 油脂类					<7>	115.76	810.32
5. 肉食及其制品					<23>	118.32	2 721.36
6. 蛋类					<6>	109.87	659.22
7. 水产品类					<7>	106.24	743.68
8. 菜类					<9>	121.11	1 089.99
9. 调味品					<2>	108.98	217.96
10. 糖类					<2>	102.45	204.9
11. 烟草类					<4>	108.43	433.72
12. 酒和饮料					<3>	112.08	336.24
13. 干鲜瓜果类					<4>	115.46	461.84
14. 糕点类					<2>	118.57	237.14
15. 奶及奶制品					<3>	120.98	362.94
16. 其他食品					<2>	117.89	235.78
17. 饮食业					<3>	122.96	368.88
二、衣着类					《9》	109.77	987.93
三、家庭设备及用品类					《6》	108.94	653.64
四、医疗保健类					《10》	113.23	1 132.3

续表

商品类别及品名	规格等级牌号	计量单位	平均价格（元）		权数	以上年为基期	
			上年	本年		指数（%）	指数×权数
（甲）	（乙）	（丙）	（1）	（2）	（3）	（4）	（5）=（4）×（3）
五、交通和通信工具类					《10》	104.19	1 041.9
六、娱乐教育与文化用品类					《11》	116.78	1 284.58
七、居住类					《13》	120.99	1 572.87
八、服务项目类					《7》	121.35	849.45

8.5.2 股票价格指数

股票价格指数即股票指数，是由证券交易所或金融服务机构编制的表明股票行市变动的一种供参考的指示数字，是通过一定方法计算出的衡量整个股票市场的股票价格总体水平和市场交易状况的一个指标。它是帮助投资者掌握股市现状和分析、判断股市变动趋势的非常重要的尺度和信号，是反映国家或地区社会经济和政治形势变动情况的最敏感指标，是整个经济的晴雨表和政治风云录。

股票价格指数因涨跌迅速，一般要求按日编制。它是以某年某月某一日股价作为基期股价（基准日股价），基准日指数通常定为 100，以后各日股价同基日股价计算出百分数，即为各日股票指数。

股票指数通常运用综合指数形式，一般以股票发行量为权数，也有以成交量为权数的。

【实例 8.6】 股票指数编制原理介绍，资料如表 8.12 所示。

表 8.12 股票指数编制资料

股票名称	发行量（股）	基准日价格（元） p_0	计算日股票价格（元/股）		
			P_1	P_2	P_3
甲	20 000	15	22	19	18
乙	30 000	11	18	19	17
丙	40 000	9	7	10	9

则第一日股票指数为：

$$\frac{\sum qp_1}{\sum qp_0} = (22 \times 20\,000 + 18 \times 30\,000 + 7 \times 40\,000) / (15 \times 20\,000 + 11 \times 30\,000 + 9 \times 40\,000)$$
$$= 1\,260\,000 / 990\,000 = 127.27\%$$

第二日股票指数为：

$$\frac{\sum qp_2}{\sum qp_0} = (19 \times 20\,000 + 19 \times 30\,000 + 10 \times 40\,000) / (15 \times 20\,000 + 11 \times 30\,000 + 9 \times 40\,000)$$
$$= 1\,350\,000 / 990\,000 = 136.36\%$$

第三日股票指数为：

$$\frac{\sum qp_3}{\sum qp_0} = (18 \times 20\,000 + 17 \times 30\,000 + 9 \times 40\,000) / (15 \times 20\,000 + 11 \times 30\,000 + 9 \times 40\,000)$$

$$= 1\,230\,000 / 990\,000 = 124.24\%$$

这说明，第一日股价上涨 27.27%，第二日比第一日上涨了 9.09 个百分点，第三日又比第二日下跌了 12.12 个百分点。

世界上几种著名的股票指数

道·琼斯股票指数

道·琼斯股票指数是世界上历史最为悠久的股票指数。它是在 1884 年由道·琼斯公司的创始人查理斯·道开始编制的。最初是根据 11 种具有代表性的铁路公司的股票，采用算术平均法进行计算编制而成，发表在查理斯·道自己编辑出版的《每日通讯》上。自 1897 年起，道·琼斯股票价格平均指数开始分成工业与运输业两大类，其中工业股票价格平均指数包括 12 种股票，运输业平均指数则包括 20 种股票，并且开始在道·琼斯公司出版的《华尔街日报》上公布。在 1929 年，道·琼斯股票价格平均指数又增加了公用事业类股票，使其所包含的股票达到 65 种，并一直延续至今。

现在的道·琼斯股票价格平均指数是以 1928 年 10 月 1 日为基期，因为这一天收盘时的道·琼斯股票价格平均数恰好约为 100 美元，所以就将其定为基准日。而以后股票价格同基期相比计算出的百分数，就成为各期的股票价格指数，所以现在的股票指数普遍用点来做单位，而股票指数每一点的涨跌就是相对于基准日的涨跌百分数。

目前，道·琼斯股票价格平均指数共分四组，第一组是工业股票价格平均指数，它由 30 种有代表性的大工商业公司的股票组成，这就是人们通常所引用的道·琼斯工业股票价格平均数。第二组是运输业股票价格平均指数，它包括 20 种有代表性的运输业公司的股票。第三组是公用事业股票价格平均指数，是由代表美国公用事业的 15 家煤气公司和电力公司的股票所组成。第四组是平均价格综合指数，它是综合前三组股票价格平均指数 65 种股票而得出的综合指数。这组综合指数虽然为优等股票提供了直接的股票市场状况，但现在通常引用的是第一组——工业股票价格平均指数。

道·琼斯股票价格平均指数是目前世界上影响最大、最有权威性的一种股票价格指数，在纽约证券交易营业时间里，每隔半小时公布一次。这一股票价格平均指数自编制以来从未间断，可以用来比较不同时期的股票行情和经济发展情况，成为反映美国股市行情变化最敏感的股票价格平均指数之一，是观察市场动态和从事股票投资的主要参考。当然，由于道·琼斯股票价格指数是一种成分股指数，它包括的公司仅占目前 2 500 多家上市公司的极少部分，而且多是热门股票，且未将近年来发展迅速的服务性行业和金融业的公司包括在内，所以它的代表性也一直受到人们的质疑和批评。

标准·普尔股票价格指数

除了道·琼斯股票价格指数外，标准·普尔股票价格指数在美国也很有影响，它是美国最大的证券研究机构（标准·普尔公司）于1923年开始编制发表股票价格指数。最初采选了230种股票，编制两种股票价格指数。到1957年，这一股票价格指数的范围扩大到500种股票，分成95种组合。其中最重要的四种组合是工业股票组、铁路股票组、公用事业股票组和500种股票混合组。从1976年7月1日开始，改为400种工业股票、20种运输业股票、40种公用事业股票和40种金融业股票。几十年来，虽然有股票更迭，但始终保持为500种。

日经道·琼斯股价指数（日经平均股价）

系由日本经济新闻社编制并公布的反映日本股票市场价格变动的股票价格平均数。该指数从1950年9月开始编制，最初根据东京证券交易所第一市场上市的225家公司的股票算出修正平均股价，当时称为“东证修正平均股价”。1975年5月1日，日本经济新闻社向道·琼斯公司买进商标，采用美国道·琼斯公司的修正法计算，这种股票指数改称“日经道·琼斯平均股价”。1985年5月1日在合同期满10年时，经两家商议，将名称改为“日经平均股价”。

按计算对象的采样数目不同，该指数分为两种：一种是日经225种平均股价，其所选样本均为在东京证券交易所第一市场上市的股票，样本选定后原则上不再更改。由于日经225种平均股价从1950年一直延续下来，因而其连续性及可比性较好，成为考察和分析日本股票市场长期演变及动态的最常用和最可靠指标。该指数的另一种是日经500种平均股价，这是从1982年1月4日起开始编制的。由于其采样包括500种股票，其代表性就相对更为广泛，但它的样本是不固定的，每年4月份要根据上市公司的经营状况、成交量和成交金额、市价总值等因素对样本进行更换。

《金融时报》股票价格指数

《金融时报》股票价格指数的全称是“伦敦《金融时报》工商业普通股股票价格指数”，是由英国《金融时报》公布发表的。该股票价格指数包括在英国工商业中挑选出来的具有代表性的30家公开挂牌的普通股股票。它以1935年7月1日作为基期，其基点为100点。该股票价格指数以能够及时显示伦敦股票市场情况而闻名于世。

香港恒生指数

香港恒生指数是香港股票市场上历史最久、影响最大的股票价格指数，由香港恒生银行于1969年11月24日开始发表。恒生股票价格指数包括从香港500多家上市公司中挑选出来的33家有代表性且经济实力雄厚的大公司股票作为成分股，分为四大类——4种金融业股票、6种公用事业股票、9种地产业股票和14种其他工商业（包括航空和酒店）股票。这些股票占香港股票市值的63.8%，因该股票指数涉及香港的各个行业，具有较强的代表性。

恒生股票价格指数的编制是以1964年7月31日为基期，基点确定为100点。其计算方法是将33种股票按每天的收盘价乘以各自的发行股数为计算日的市值，再与基期的市值相比较，乘以100就得出当天的股票价格指数。开始时，恒生指数只计算一个总指数，为方便投资者获得更多信息，自1984年1月13日起，增算金融业、地产业、工商业和公用事业4个分类指数。

由于恒生股票价格指数所选择的基期适当，因此，不论股票市场狂升或猛跌，还是处于正常交易水平，恒生股票价格指数基本上能反映整个股市的活动情况。

8.6 Excel在统计指数分析中的应用

8.6.1 利用Excel进行指数计算

在 Excel 中，没有专门用于指数计算和因素分析的工具与统计函数，一般利用公式和一般函数进行相应的分析。

1. 综合指数的计算

以【实例 8.7】作为范例，说明在 Excel 中计算综合指数的操作方法。

【实例 8.7】 根据某企业三种产品单位成本和产量资料计算单位成本指数。

原始数据及各步计算数据如图 8.1 所示，计算综合指数具体步骤如下。



	A	B	C	D	E	F	G	H	I	J
	产品名称	计量单位	基期 q_0	报告期 q_1	基期 p_0	报告期 p_1	$p_0 \cdot q_1$	$p_1 \cdot q_1$		
1										
2	甲	万件	80	120	24	20	2880	2400		
3	乙	万台	60	60	18	18	1080	1080		
4	丙	万套	50	30	15	19	450	570		
5							4410	4050		
6										
7										
8										
9	单位产品指数									
10										
11										
12										
13										
14										

图 8.1 综合指数计算数据

第一步：计算每一个“ $p_0 \cdot q_1$ ”。在 G3 单元格中输入“=D3*E3”，回车得到结果；然后使用填充柄功能按住鼠标左键向下拖，至 G5 单元格，放开鼠标，可得 G4~G5 结果，即自动填入 G4~G5 的数值。

第二步：计算每一个“ $p_1 \cdot q_1$ ”。在 H3 单元格中输入“=D3*F3”，回车得到结果；然后使用填充柄功能按住鼠标左键向下拖，至 H5 单元格，放开鼠标，可得 H4~H5 结果，即自动填入 H4~H5 的数值。

第三步：计算“ $\sum p_0 \cdot q_1$ ”和“ $\sum p_1 \cdot q_1$ ”。选定 G3: G5 区域，单击工具栏上的“ Σ ”按钮，在 G6 单元格出现该列的求和值。选定 H3: H5 区域，单击工具栏上的“ Σ ”按钮，在 H6 单元格出现该列的求和值。

第四步：计算单位成本指数“ $\sum p_1 \cdot q_1 / \sum p_0 \cdot q_1$ ”。单击任一空单元格，本例中为 C9，输入“=H6/G6”，即可得单位成本指数 91.836 7%。

数量指标综合指数与质量指标综合指数的计算方法类似，这里就不再一一赘述。

2. 平均指数的计算

以【实例 8.8】作为范例，说明在 Excel 中计算平均指数的操作方法。

【实例 8.8】 根据某企业三种产品单位成本和报告期总成本资料利用加权调和平均数指数计算单位成本指数。

原始数据及各步计算数据如图 8.2 所示，计算平均数指数具体步骤如下。

	A	B	C	D	E	F	G	H	I
1			单位成本 (元)	报告期总成本					
2	产品名称	计量单位	基期 p_0	报告期 p_1	$p_1 \cdot q_1$	$k=p_1/p_0$	$(p_1 \cdot q_1)/k$		
3	甲	万件	24	20	2400	0.833333	2880		
4	乙	万台	18	18	1080	1	1080		
5	丙	万套	15	19	570	1.266667	450		
6					4050		4410		
7									
8									
9	单位产品指数		0.918367						
10									
11									
12									
13									
14									

图 8.2 平均指数计算数据

第一步：计算每一个个体指数“ $k=p_1/p_0$ ”：在 F3 单元格中输入“=D3/C3”，回车得到结果；然后使用填充柄功能按住鼠标左键向下拖，至 F5 单元格，放开鼠标，可得 F4~F5 结果，即自动填入 F4~F5 个体指数的数值。

第二步：计算每一个“ $p_1 \cdot q_1/k$ ”。在 G3 单元格中输入“=E3/F3”，回车得到结果；然后使用填充柄功能按住鼠标左键向下拖，至 G5 单元格，放开鼠标，可得 G4~G5 结果，即自动填入 G4~G5 的数值。

第三步：计算“ $\sum p_1 \cdot q_1$ ”和“ $\sum p_1 \cdot q_1/k$ ”。选定 E3：E5 区域，单击工具栏上的“ \sum ”按钮，在 E6 单元格出现该列的求和值。选定 G3：G5 区域，单击工具栏上的“ \sum ”按钮，在 G6 单元格出现该列的求和值。

第四步：计算单位成本指数“ $\sum p_1 \cdot q_1 / \sum p_1 \cdot q_1/k$ ”。单击任一空单元格，本例中为 C9，输入“=E6/G6”，即可得单位成本指数 91.836 7%。

加权算术平均数指数与加权调和平均指数的计算方法类似，这里就不再举例说明。

8.6.2 利用Excel进行因素分析

利用【实例 8.7】的数据，以【实例 8.9】作为范例，说明在 Excel 中进行因素分析的操作方法。

【实例 8.9】 根据某企业三种产品单位成本和产量资料进行因素分析。

原始数据及各步计算数据如图 8.3 所示, 进行因素分析的具体步骤如下。

	A	B	C	D	E	F	G	H	I	J	K
	产品名称	计量单位	基期 _{q0}	报告期 _{q1}	单位成本 _{p0}	报告期 _{p1}	$P_0 \cdot q_1$	$P_1 \cdot q_1$	$P_0 \cdot q_0$		
1	甲	万件	80	120	24	20	2880	2400	1920		
2	乙	万台	60	60	18	18	1080	1080	1080		
3	丙	万套	50	30	15	19	450	570	750		
4							4410	4050	3750		
5	单位产品指数		0.918367		由于单位成本变动对总成本的影响指数				-360		
6	产品产量指数		1.176		由于产量变动对总成本的影响指数				660		
7	总成本指数		1.08		总成本变动				300		

图 8.3 因素分析数据

第一步：计算每一个“ p_0*q_1 ”和“ $\sum p_0*q_1$ ”：在 G3 单元格中输入“=D3*E3”，回车得到结果；然后使用填充柄功能按住鼠标左键向下拖，至 G5 单元格，放开鼠标，可得 G4~G5 结果。选定 G3:G5 区域，单击工具栏上的“ \sum ”按钮，在 G6 单元格出现该列的求和值。

第二步：计算每一个“ p_1*q_1 ”和“ $\sum p_1*q_1$ ”：在 H3 单元格中输入“=D3*F3”，回车得到结果；然后使用填充柄功能按住鼠标左键向下拖，至 H5 单元格，放开鼠标，可得 H4~H5 结果。选定 H3: H5 区域，单击工具栏上的“ \sum ”按钮，在 H6 单元格出现该列的求和值。

第三步：计算每一个“ $p_0 \cdot q_0$ ”和“ $\sum p_0 \cdot q_0$ ”。在 I3 单元格中输入“=C3*E3”，回车得到结果；然后使用填充柄功能按住鼠标左键向下拖，至 I5 单元格，放开鼠标，可得 I4~I5 结果。选定 I3: I5 区域，单击工具栏上的“ \sum ”按钮，在 I6 单元格出现该列的求和值。

第四步：计算单位成本指数、产量指数和总成本指数。在 C9 单元格中输入“=H6/G6”，求得单位成本指数；在 C10 单元格中输入“=G6/I6”，求得产量指数；在 C11 单元格中输入“=H6/I6”，求得总成本指数。

第五步：计算由于单位成本变动对总成本的影响数、由于产量变动对总成本的影响数和总成本变动总数。在 I9 单元格中输入“=H6-G6”，求得由于单位成本变动对总成本的影响数；在 I10 单元格中输入“=G6-I6”，求得由于产量变动对总成本的影响数；在 I11 中输入“=H6-I6”，求得总成本变动总数。



统计术语

数量指数 index number of quantity

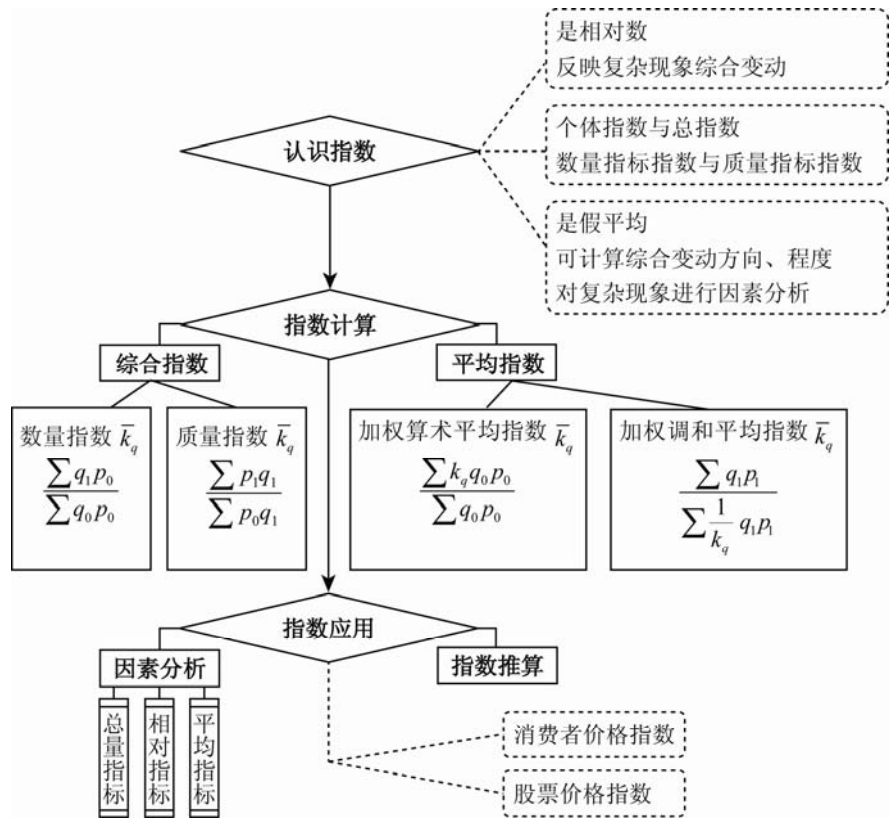
拉氏 laspeyres

价格指数 price index

报告期 the period of interest



重点知识梳理



习题与实践训练

一、判断题

1. 分析复杂现象总体的数量变动，只能采用综合指数的方法。 ()
2. 在实际应用中，计算价格综合指数，需要采用基期数量指标为同度量因素。 ()
3. 分析复杂现象总体的数量变动时，若研究的是数量指标的变动，则选择的同度量因素是数量指标。 ()
4. 某企业的某种产品单位成本 2016 年 3 月与去年同期相比上升了 8%，产量下降了 8%，则总成本没升也没降。 ()
5. 可变构成指数=固定构成指数+结构影响指数。 ()
6. 设 p 表示价格， q 表示销售量，则 $\sum p_0 q_1 - \sum p_0 q_0$ 表示由于商品价格的变动对商品总销售额的影响。 ()
7. 从指数化指标的性质来看，单位成本指数是数量指标指数。 ()
8. 总指数有两种计算形式，即个体指数和综合指数。 ()
9. 商品价格上涨 5%，商品销售量增长 3%，则商品销售额增长 15%。 ()

10. 计算总指数时，为了解决总体各要素的量不能直接相加而使用的一个媒介因素，叫同度量因素。（ ）

二、单选题

- 统计指数按指数化指标的性质不同，可分为（ ）。
A. 总指数和个体指数
B. 数量指标指数和质量指标指数
C. 平均数指数和平均指标指数
D. 综合指数和平均数指数
- 下列指数中（ ）是数量指标指数。
A. 商品物价指数
B. 平均工资指数
C. 单位成本指数
D. 销售量指数
- 总指数的两种计算基本形式是（ ）。
A. 个体指数和综合指数
B. 算术平均指数和调和平均指数
C. 综合指数和平均指数
D. 可变构成指数、固定构成指数和结构影响指数
- 如果消费价格指数上涨 20%，则现在 1 元钱（ ）。
A. 只值原来的 0.8 元钱
B. 只值原来的 0.83 元钱
C. 与原来的 1 元钱等值
D. 无法与原来的钱比较
- 价格下降后，花同样多的钱可多购买基期商品的 10%，则物价指数为（ ）。
A. 90%
B. 90.9%
C. 110%
D. 111.1%
- 总销售量增加，销售额持平，则物价指数（ ）。
A. 降低
B. 增长
C. 不变
D. 无法确定
- 已知某商场的商品销售量指数为 105%，由于销售量增长而增加的销售额为 10 万元；又知道销售价格指数为 110%，由于价格上涨而增加的销售额应为（ ）。
A. 30 万元
B. 21 万元
C. 20 万元
D. 18 万元
- 当我们研究各级别工人工资的变动对全体工人平均工资变动的影响程度时，应计算（ ）。
A. 结构影响指数
B. 可变构成指数
C. 固定构成指数
D. 加权算术平均指数
- 某商店今年与去年相比，商品销售量指数下降了 10%，销售价格指数上涨 10%，则商品销售额将（ ）。
A. 不变
B. 上升
C. 下降
D. 可能上升也可能下降
- 某企业的职工工资水平比上年提高了 5%，职工人数增长了 2%，则该企业工资总额增长（ ）。
A. 11%
B. 10%
C. 7.1%
D. 7%

三、多项选择题

- 下列指数中（ ）是质量指标指数。
A. 商品价格指数
B. 平均工资指数

- C. 单位成本指数 D. 职工人数指数
E. 商品销售量指数
2. 同度量因素的作用有 ()。
- A. 平均作用 B. 比较作用 C. 权数作用
D. 稳定作用 E. 同度量作用
3. 某地区商业企业职工去年劳动生产率指数为 132%，这是 ()。
- A. 个体指数 B. 总指数 C. 平均指标指数
D. 数量指标指数 E. 质量指标指数
4. 平均指标变动因素分析的指数体系中包括的指数有 ()。
- A. 可变构成指数 B. 加权算术平均指数 C. 固定构成指数
D. 结构影响指数 E. 加权调和平均指数
5. 某企业基期产值为 100 万元，报告期产值比基期增长 14%，又知以基期价格计算的报告期假定产值为 112 万元，则经过计算可知 ()。
- A. 由于产量变化使产值增加 12 万元 B. 产量增加 12%
C. 由于产量变化使产值增加 20 万元 D. 价格增加 12%
E. 由于价格变化使产值增加 2 万元
6. 下列情况中，属于广义指数概念的有 ()。
- A. 不同空间同类指标之比
B. 同类指标实际与计划之比
C. 一种指标与另一种有密切联系的指标之比
D. 同一总体的部分指标与总量指标之比
E. 同一总体的部分指标与另一部分指标之比
7. 下列属于数量指标指数的有 ()。
- A. 工业生产指数 B. 劳动生产率指数 C. 职工人数指数
D. 产品产量指数 E. 产品单位成本指数
8. 编制综合指数的原则是 ()。
- A. 质量指标指数以报告期的数量指标作为同度量因素
B. 质量指标指数以基期的数量指标作为同度量因素
C. 数量指标指数以基期的数量指标作为同度量因素
D. 数量指标指数以基期的质量指标作为同度量因素
E. 数量指标指数以固定时期质量指标作为同度量因素
9. 对某商店某时期商品销售额变动情况分析，其指数体系包括 ()。
- A. 销售量指数 B. 销售价格指数 C. 总平均价格指数
D. 销售额指数 E. 个体指数
10. 进行平均指标变动的因素分析应编制的指数有 ()。
- A. 算术平均数指数 B. 调和平均数指数 C. 可变构成指数
D. 固定构成指数 E. 结构影响指数

四、填空题

1. 在平均数指数的基本形式中，编制数量指标指数的常用形式是_____，编制质量指标指数的常用形式是_____。
2. 某商店商品销售额报告期和基期相同，报告期商品价格比基期提高了 10%，那么，报告期商品销售量比基期_____。
3. 2015 年与 2014 年相比，某机关的职工工资水平提高了 5%，职工人数增加了 1%，则职工工资总额增长了_____。
4. 三个或三个以上在经济上有联系，数量上保持一定对等关系的指数构成的整体体系是_____体系。
5. 在编制数量指标指数时，一般以_____作为同度量因素，把它固定在_____期；在编制质量指标指数时，一般以_____作为同度量因素，把它固定在_____期。
6. 平均指数是_____的加权平均数。常用的基本形式有两种，一种是加权_____平均数指数，另一种是加权_____平均数指数。
7. 单位成本指数 = $\frac{\sum z_1 q_1}{\sum z_0 q_1} = 97\%$ ， $\sum z_1 q_1 - \sum z_0 q_1 = -2\,200$ （元），则相对数的含义是_____，绝对数的含义是_____。
8. 用同样多的人民币，今年比去年少购买商品 5%，则物价指数是_____。
9. 职工平均工资增长了 3.5%，固定构成指数是 15%，则职工人数结构影响指数下降或提高了_____。
10. 某空调厂 2016 年 5 月与 2015 年 5 月相比，空调产量增长了 16%，总成本上升了 17.2%，则单位成本上升或下降_____。

五、应用能力训练题

1. 区分下表中各指数的种类。

统计指数	个体指数	总指数	数量指标指数	质量指标指数
某一种商品单位成本指数				
二十种商品的价格指数				
企业职工的平均工资指数				
某型号手机的销售量指数				
全国消费品零售价格指数				

2. 某商场三种商品的实际销售额资料如下。

商品名称	计量单位	实际销售额（万元）		8 月销售量比 7 月销售量增长的百分比（%）
		7 月	8 月	
甲	件	1 000	1 200	5
乙	袋	2 000	2 000	0
丙	台	3 000	3 200	10

计算分析下列问题：

- (1) 各种商品的个体价格指数和个体销售量指数。
- (2) 三种商品的物价总指数和销售量总指数。
- (3) 三种商品的销售额总指数。
- (4) 从相对数和绝对数上分析销售额总指数与价格总指数、销售量总指数之间的经济关系。

3. 某大学生自主创业，经营绿色早餐摊点，主要销售小笼包和粥两种传统食品。销售资料如下。

商品名称	计量单位	销售额（元）		4 月销售量比 3 月销售量增长的百分比（%）
		3 月	4 月	
小笼包	屉	15 000	20 000	-2%
粥	碗	2 850	3 600	2%
合计	—	17 850	23 600	—

求该摊点总的说来 4 月与 3 月比较销售量的变动情况。

4. 某大学生自主创业，经营绿色早餐摊点，主要销售小笼包和粥两种传统食品。销售资料如下。

商品名称	计量单位	销售额（元）		4 月价格比 3 月价格增长的百分比（%）
		3 月	4 月	
小笼包	屉	15 000	20 000	20%
粥	碗	2 850	3 600	10%
合计	—	17 850	23 600	—

求该摊点总的说来 4 月与 3 月比较价格的变动情况。

5. 某公司 2015 年第二季度和第一季度的产品产值和单价资料如下。

产品名称	产品产值（万元）		第二季度价格是第一季度的百分比（%）
	第一季度	第二季度	
甲	400	600	105
乙	550	450	115
丙	200	350	98
合计	1 150	1 400	—

要求：分析产品单价和产量的变动对产值的变动影响。

6. 某商店商品的价格和销售量资料如下。

商品名称	计量单位	价格（元）		销售量	
		基期	报告期	基期	报告期
甲	双	48	45	300	400
乙	件	30	32	400	500

要求：

- (1) 计算某商店几种商品的价格总指数和销售量总指数。
- (2) 指出在总销售额增长绝对值中，有多少绝对数是受价格因素影响的，有多少绝对数是

受销售量因素影响的？

7. 某企业某种产品生产总成本 2015 年为 12.9 万元，比 2014 年多 9 000 元，单位产品成本比 2014 年降低 3%，试确定：

- (1) 生产总成本指数。
- (2) 产品产量指数。
- (3) 由于单位成本降低而节约的绝对额。

8. 某地区 2015 年社会商品零售额为 2 570 亿元，比 2014 年增长 9.4%，剔除零售物价上升的因素，社会商品零售额实际增长 7.3%。试计算：

- (1) 2015 年与 2014 年比较，零售物价上升多少？
- (2) 计算分析 2015 年比 2014 年社会商品零售额增长的数量和原因。

9. 某管理局所属三个工厂生产同种产品，它们的单位产品成本和产量资料如下。

厂 名	产量（件）		每件成本（元）	
	2014 年	2015 年	2014 年	2015 年
一厂	1 600	2 400	10.0	9.0
二厂	1 800	2 400	10.4	9.2
三厂	2 400	1 600	9.6	9.6

根据上述资料：

- (1) 分别计算三个工厂生产各种产品的 2014 年和 2015 年总平均成本；
- (2) 计算平均成本指数，分析由于平均成本下降而节约的总成本金额；
- (3) 在平均成本的总变动中，分析各工厂成本水平变动及各工厂产量结构变动的影响程度相对数和绝对值。

10. 某公司商品的价格和销售量资料如下。

商品名称	单位	商品价格（元）		销售量	
		基期	报告期	基期	报告期
甲	件	2.5	2	1 000	1 050
乙	吨	65	60	180	200
丙	个	1.5	1.2	1 500	1 850

要求利用 Excel 软件完成下列计算：

- (1) 计算三种商品价格指数和销售量指数。
- (2) 分析说明销售额总变动中各因素变动的影响程度和影响绝对数。



本章案例

美媒：全球幸福排名瑞士夺冠，北欧国家表现最佳

你认为自己的国家“幸福”吗？《2015 年全球幸福指数报告》根据包括经济与社会福利在内的一组变量，对 158 个国家和地区进行排名，一些结果可能会让你大吃一惊。

据美国之音电台网站 5 月 19 日报道,如同一些热门的单曲一样,幸福是超越国界的。不过,各国幸福的原因千差万别。由“可持续发展解决方案网络”(the Sustainable Development Solutions Network)公布的《2015 年全球幸福报告》(the World Happiness Report)利用经济学、心理学、健康与公共政策专家来界定和衡量全球各地的幸福指数。

在报告中,瑞士取代丹麦,摘得桂冠。而丹麦则名列第三。北欧国家继续整体表现最佳,而美国则名列第 15 位,排在巴西之前,但位居邻国墨西哥之后。

克莱尔·布杰是报告的编辑与协调员。她说,高收入是排在首位通用变量之一,但高收入并不能自动转换成幸福。

她说:“一些基本的收入水平的确很重要,因为你必须要能够满足自己的基本需求,你必须能养活自己;但在进入了一定的收入门槛之后,收入水平对于是否幸福来说,其实就并不重要了。”

排名领先的国家通常腐败程度较低而社会支持水平强。换言之,在有需要的时候能够有可以依靠的人。布杰说,名列第二的冰岛与名列 102 位的希腊之间最大的区别就是社会支持度。两国都受金融危机影响,但应对却不同。

相比之下,全球排名最低的国家 and 地区往往与极端贫困相关。最不幸的 20 个国家多数在撒哈拉以南非洲地区,但有两个例外。

布杰说:“现在我们看到叙利亚已经跌至名单的最后,中非共和国也排在最后,还有阿富汗。所以说,的确都是一些非常贫穷,以及冲突和暴力严重的国家。”

在亚洲,日本位居第 46 名,中国排名 84。

报告显示,其他确定幸福程度的关键变量包括:健康预期寿命、人生选择自由及乐善好施的程度。

(资料来源:参考消息 2015-05-19)

相关分析与回归分析

学习要点

- 相关关系的概念与种类。
- 相关系数的意义与计算。
- 相关分析与回归分析的特点。
- 一元线性回归方程的确定方法及估计标准误差的计算。
- 回归系数及估计标准误差的意义。

在日常生活中，许多现象之间存在相互联系、相互制约的关系。例如，商品的销售额与商品的广告费之间有密切关系，所以商家往往不惜重金做广告，以期得到高销售额回报；公安机关根据案发现场遗留脚印的大小，可以推测出犯罪嫌疑人的性别和身高；某地区的人均消费水平与该地的人均国内生产总值（GDP）常存在着高度相关的关系，等等。那么，我们如何采用科学方法判断现象之间是否存在关系？怎样用量化数值度量关系的密切程度？对关系非常密切的现象（即变量）怎样更进一步表达它们之间的关系？解决这些问题需要用到本章的统计相关分析和回归分析工具。

9.1 相关分析

9.1.1 相关关系的概念

相关关系（correlation）是指现象间的非确定性的数量上的依存关系。它有两个特点：一是现象之间确实存在数量上的依存关系；二是数量依存关系的值是不确定的。例如，人的身高和体重之间确实存在着“身高越高，体重越大”的数量关系，但是，身高和体重却不是一一对应的。身高 170cm 的人，相应的体重并不完全一样；同样，体重 65kg 的人，会对应着不同的身高。相关关系的这种特点，决定了它与函数关系的区别。

函数关系是指现象间存在的确定性的数量依存关系。在这种关系中，某个现象的数值发生变化，都有另一现象的确定值与它对应。这种关系可用数学函数式反映出来。例如，给定一个圆的半径，就有唯一确定的圆的面积和它对应，面积是半径的函数。在社会经济中，产品生产费用（总成本）=产品产量×单位产品成本，当单位产品成本不变时，产品产量发生变化，就有

一个确定的总成本与它对应，总成本是产品产量的函数。

函数关系与相关关系既有区别又有联系。由于观察和实验中的误差，函数关系往往通过相关关系表现出来。而当对现象之间的内在联系和规律性了解得更加清楚的时候，相关关系有可能转化为函数关系。一般说来，在社会经济领域里，函数关系反映了现象之间的理想状态，而相关关系则反映了现象之间的现实状态。在对具有相关关系的现象进行分析时，必须利用相应的函数关系数学表达式来表明现象之间的相关方程式。相关关系是相关分析（correlation analysis）的研究对象，函数关系是相关分析的工具。

9.1.2 相关关系的种类

现象的相关关系可以按不同的标志进行区分。

1. 按相关程度区分

按相关程度不同，相关关系可分为完全相关、不相关和不完全相关。

两种现象之间，若当一个现象的数量确定时，另一现象的数量也随之确定，则这两种现象间的关系称为完全相关，其关系如图 9.1 所示。例如，中国银行的 1 年期人民币居民储蓄存款利率为 2.5%，存入的本金用 x 表示，到期的本利和用 y 表示，则 $y=x+2.5\%x$ 。这里 y 与 x 表现为一种线性函数关系，也就是说，函数关系是相关关系的一个特例。

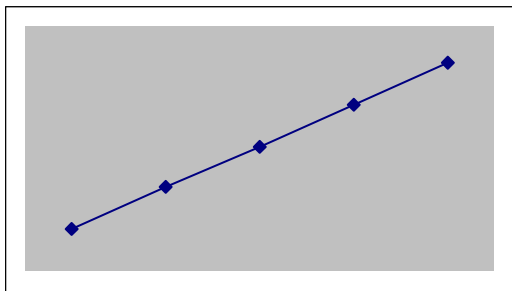


图 9.1 完全相关

如果两个现象之间的数量变化相互独立，彼此互不影响，则这两种现象间的关系称为不相关，其关系如图 9.2 所示。例如，学生的学习成绩与身高之间不存在相关关系。

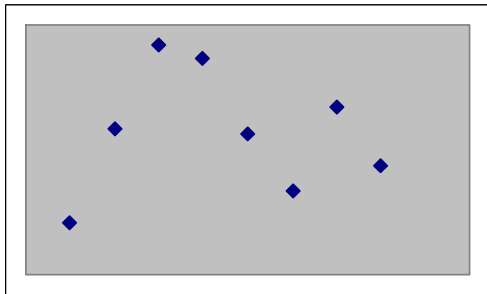


图 9.2 不相关

如果两个现象之间的关系介于完全相关和不相关之间，则称为不完全相关。例如，妇女的结婚年龄与受教育程度之间是不完全相关关系。现实经济生活中大多数相关关系都属于不完全相关，这是统计分析的主要研究对象，其关系如图 9.3 所示。

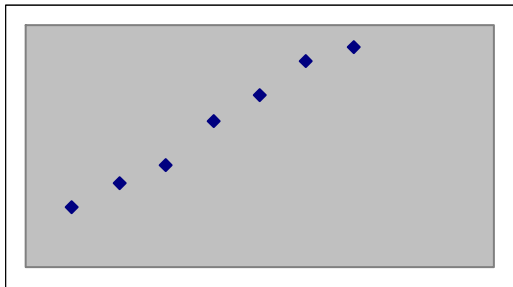


图 9.3 不完全相关

2. 按相关方向区分

按相关方向不同，相关关系可分为正相关和负相关。

相关变量的变化方向一致，即一个变量数值增大（或减少），另一个变量的数值也相应地增大（或减少），这种相关关系称为正相关，如图 9.3 所示。如居民的收入与居民的储蓄额的关系，属于正相关。

若相关变量的变化方向相反，即一个变量数值增大（或减少），另一个变量的数值相应地减少（或增大），这种相关关系称为负相关，如图 9.4 所示。例如，家庭收入增加，恩格尔系数下降，它们之间属于负相关关系。

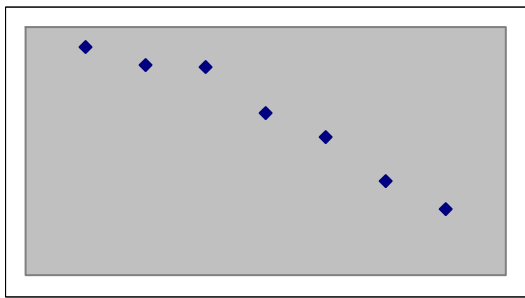


图 9.4 负相关

3. 按相关表现形式区分

按相关的表现形式不同，相关关系可分为线性相关和非线性相关。

对两个具有相关关系的现象进行实际调查，能获得一系列成对的数据。每对数据在平面直角坐标系中确定一个点，如果这些点大致分布在一条直线周围，则这两种现象就构成线性相关，如图 9.3 和图 9.4 所示。例如，在一定范围内，出租汽车费用与行驶里程呈线性相关关系。若这些点的分布并不表现为直线关系，而是近似于某种曲线关系，则这种关系就称为非线性相关，如图 9.5 所示。而从人的生命过程来看，年龄与医疗费支出呈非线性相关。

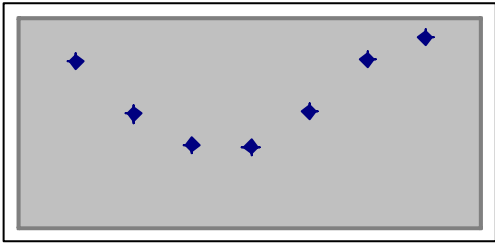


图 9.5 非线性相关

4. 按相关变量多少区分

按相关的变量多少不同，相关关系可分为单相关和复相关。

两个变量之间的相关关系称为单相关，这种相关关系中只涉及一个自变量和一个因变量，如图 9.1 至图 9.5 所示。例如，只研究居民的收入和居民的消费的关系就是单相关。

3 个或 3 个以上变量的相关关系称为复相关。这种相关关系中涉及两个或两个以上自变量和一个因变量，如冰激凌的销售量与价格、气温、居民收入及居民年龄构成等之间的相关关系就是复相关。

统计学中对相关关系进行分析的主要目的，是要对现象之间的相互关系的密切程度和变化规律进行分析，取得具体数量上的认识，并进一步确定出相互关系的模式，以便于进行统计预测和推算。根据对现象分析研究层次不同，相关分析主要内容有狭义与广义之分。

狭义地讲，相关分析主要内容有：

- (1) 判断现象之间是否存在相关关系（通过编制相关表和绘制相关图）。
- (2) 确定相关关系的密切程度和方向（通过计算相关系数）。

广义地说，相关分析主要内容除了包括狭义相关分析内容外，还包括：

- (1) 确定相关关系的数学表达式（通过建立回归方程）。
- (2) 确定因变量估计值误差的程度（通过计算估计标准误差）。

下面逐项进行研究。

9.1.3 相关图表

相关分析的起点是定性分析，按照认识的逻辑顺序分析并判断现象之间是否存在相关关系，在此基础上，进一步进行定量分析，编制相关表，绘制相关图，利用相关图表，判断相关关系的形式，为相关分析奠定基础。

1. 相关表

相关表（correlation table）是一种反映变量之间相关关系的统计表。将其中一个变量按其取值大小顺序排列，再将其相关的另一变量数值对应列出来，便可形成简单相关表。通过相关

表，可初步看出相关关系的形式、密切程度和相关方向。现举例说明。

【实例 9.1】 对我国部分省市人均 GDP 及人均消费水平进行调查，得原始资料如表 9.1 所示，再通过相关表，判断两者的相关关系。

表 9.1 2013 年我国部分省市人均 GDP 及人均消费水平原始资料 单位：元

地区	北京	浙江	上海	黑龙江	山东	贵州	陕西
人均 GDP	93 213	68 462	90 092	37 509	56 323	22 922	42 692
人均消费水平	33 337	24 771	39 223	12 978	16 728	9 541	13 206

(资料来源：中国统计年鉴 2014)

解：根据以上原始资料，将人均 GDP 按从小到大顺序排列，可编制简单相关表，如表 9.2 所示。

表 9.2 2013 年我国部分省市人均 GDP 及人均消费水平相关表 单位：元

地区	贵州	黑龙江	陕西	山东	浙江	上海	北京
人均 GDP	22 922	37 509	42 692	56 323	68 462	90 092	93 213
人均消费水平	9 541	12 978	13 206	16 728	24 771	39 223	33 337

从表 9.2 可以看出，随着人均 GDP 的提高，人均消费水平也有相应提高的趋势，两者之间存在着明显的正相关关系。

2. 相关图

相关图（correlation diagram）又称散点图（scatter plot），它是把相关表中的原始对应数据在平面直角坐标系中以坐标点描绘出来。以横轴表示自变量（dependent variable），纵轴表示因变量（independent variable），通过标出的每对变量值的坐标点或散布点，观察其分布状况。例如，可将表 9.1 中所列的资料绘制成相关图，如图 9.6 所示。

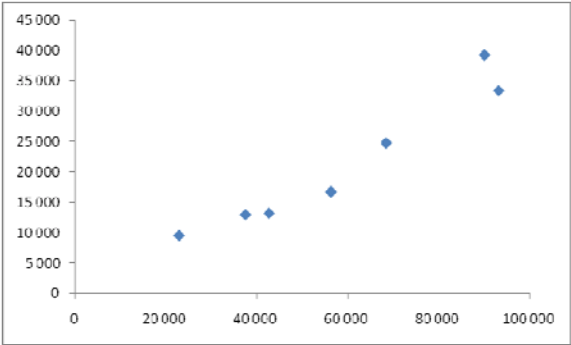


图 9.6 人均 GDP 及人均消费水平相关图

从图 9.6 中关于人均 GDP 及人均消费水平的 7 个散布点来看，人均消费水平随着人均 GDP 的增大而增大，随之减小而减小，并且散布点的分布近似地表现为一条直线。由此可以判断，人均消费水平与人均 GDP 这两个变量之间存在着直线正相关关系。

但是，相关表和相关图只能大体上反映现象之间的相关关系，为了更为精确地反映关系类

型的特征, 我们还应该进一步用统计分析指标——相关系数来表明相关的密切程度。按照相关变量的个数和分析问题的角度不同, 相关系数可以分为简单相关系数、偏相关系数和复相关系数。在本章中, 我们只讨论简单相关系数, 它适用于影响因变量的因素只有一个自变量的情况。

9.1.4 相关系数

1. 相关系数的意义

相关系数 (correlation coefficient) 是在直线相关条件下, 说明两个变量之间相关关系密切程度和方向的统计分析指标, 通常用 r 表示。其表现形式为相对数, 它不受变量值水平和计量单位的影响。

相关系数的取值有 1 个范围, $-1 \leq r \leq 1$; 有两个方向, $r > 0$ 表示正相关, $r < 0$ 表示负相关; 有 3 个特殊点, $r=1$ 表示完全正线性相关, $r=-1$ 表示完全负线性相关, $r=0$ 表示完全不线性相关; 有 4 个相关密切程度等级, $0 < |r| < 0.3$ 为微相关, $0.3 \leq |r| < 0.5$ 为低度相关, $0.5 \leq |r| < 0.8$ 为显著相关, $0.8 \leq |r| < 1$ 为高度相关。

采用上述标准进行判断时, 需要注意 r 只表示 x 和 y 的直线相关密切程度, 当 $|r|$ 很小甚至等于 0 时, 并不一定表示 x 与 y 之间就不存在其他非直线类型的关系。另外, 要注意应该有较多原始数据计算相关系数。当原始资料较少时, 为了判断的准确, 最好对相关系数的显著性进行检验。关于检验方法, 本书不进行讨论。

2. 相关系数的计算

在直线相关的条件下, 相关系数的定义公式是通过自变量和因变量的各个离差的乘积来表明相关关系的密切程度的, 所以用这种方法计算的相关系数叫积差法。

相关系数的基本公式如下:

$$r = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y} \quad (9-1)$$

式中: $\sigma_{xy}^2 = \frac{\sum (x - \bar{x})(y - \bar{y})}{n}$ 称为协方差;

$\sigma_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$ 是 x 的标准差;

$\sigma_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}}$ 是 y 的标准差。

所以相关系数可表现为如下形式:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \cdot \sum (y - \bar{y})^2}} \quad (9-2)$$

这是计算相关系数的基本公式，但在实际工作中利用它来计算相关系数相当烦琐，经代数演算，可化简为较简单的、能直接用相关表中的原始数据计算的计算公式。计算公式为：

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \cdot \sqrt{n \sum y^2 - (\sum y)^2}} \tag{9-3}$$

式中：*n*——相关表中数据项数。

【实例 9.2】 根据表 9.1 的资料，已知人均 GDP 与人均消费水平之间为直线相关关系，计算人均 GDP 与人均消费水平的相关系数。

解：根据表 9.1 的数据，可以编制人均 GDP 与人均消费水平的相关系数，如表 9.3 所示。

编 号	人均 GDP <i>x</i>	人均消费水平 <i>y</i>	<i>x</i> ²	<i>y</i> ²	<i>xy</i>
1	22 922	9 541	525 418 084	91 030 681	218 698 802
2	37 509	12 978	1 406 925 081	168 428 484	486 791 802
3	42 692	13 206	1 822 606 864	174 398 436	563 790 552
4	56 323	16 728	3 172 280 329	279 825 984	942 171 144
5	68 462	24 771	4 687 045 444	613 602 441	1 695 872 202
6	90 092	39 223	8 116 568 464	1 538 443 729	3 533 678 516
7	93 213	33 337	8 688 663 369	1 111 355 569	3 107 441 781
合计	411 213	149 784	28 419 507 635	3 977 085 324	10 548 444 799

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \cdot \sqrt{n \sum y^2 - (\sum y)^2}} = 0.964\ 3$$

计算结果表明，人均 GDP 与人均消费水平之间存在高度正直线相关关系。

9.2 回归分析

9.2.1 回归分析的意义

回归（regression）这个术语是由英国著名统计学家 Francis Galton（1822—1911）在 19 世纪末期在研究孩子及他们的父母的身高时提出来的。Galton 对 1 078 对父子身高关系进行了研究，假设父亲的身高是影响的因素，记为 *x*（单位：英寸，1 英寸≈2.54 厘米），其儿子的身高是影响的结果，记为 *y*（单位：英寸）。经过对资料进行整理、分组后，确定出那些资料中，大约存在如下的线性关系：

$$y_c=33.73+0.516x$$

式中：*y_c*表示假设父亲身高为 *x* 时其儿子的身高的估计值。

这就是著名的回归直线方程的雏形。之所以称为回归直线方程，首先它是直线方程，斜率 0.516 表明，父亲身高每增加 1 英寸时，其儿子的身高平均增加大约 0.5 英寸。所谓回归是指，虽然该方程反映 y_c 与 x 的关系是正比的关系，但是，当 x 较小（父亲身高较矮）时， y_c （儿子的身高）可能比 x 大；而当 x 较大时， y_c 可能比 x 小，有“回归”之意。Galton 把这种孩子的身高向中间值靠近的趋势称为一种回归效应，并把“回归”作为统计概念加以应用。但现今统计学的“回归”概念已不是指原来生物学上的特殊规律性，而是指变量之间的依存关系。

回归分析 (regression analysis) 是指研究一个或几个变量的变动对另一个变量的变动影响程度的方法。根据资料，建立变量之间的回归数学模型，即建立相应的数学表达式，通过给定的自变量的数值估计因变量的可能值，这种分析就叫回归分析，所建立的数学模型称回归模型。回归模型有多种形式，按自变量个数分，有一元回归（即简单回归，simple regression）和多元回归（即复回归，multiple regression）；按回归线形状分，有线性回归（即直线回归，linear regression）和非线性回归（即曲线回归，nonlinear regression）。其中，线性回归是基本的，本章只介绍一元线性回归，即简单线性回归分析方法。

9.2.2 回归分析的特点

相关分析是回归分析的基础和前提，回归分析是相关分析的深入和继续。

回归分析与相关分析相比具有以下特点：

(1) 在相关分析中，两变量是随机变量；而回归分析中，因变量是随机变量，自变量是给定的数值。

(2) 相关分析只能计算出一个反映两个变量间相关密切程度的相关系数，且两个变量是对等关系，变量 x 和 y 位置互换不影响相关系数的数值。而回归分析所研究的两个变量不是对等关系，有时可根据研究目的不同建立两个不同的回归方程，一个是以 x 为自变量， y 为因变量的“ y 对 x 的回归方程”；另一个以 y 为自变量， x 为因变量的“ x 对 y 的回归方程”。

(3) 相关分析计算的相关系数是一个绝对值在 0 与 1 之间的抽象系数，其数值大小反映变量之间相关关系的程度；而回归分析建立的回归方程反映的是变量值之间的具体变动关系，不是抽象的系数。根据回归方程，利用自变量的给定值，可以估计或推算因变量的数值。

9.2.3 一元线性回归方程

一元线性回归方程 (regression equation)，亦称直线方程，是分析一个自变量 x 与一个因变量 y 之间线性关系的数学方程。方程的基本形式是：

$$y_c = a + bx \quad (9-4)$$

a 和 b 称为回归方程中的两个待定参数，是需要根据相关表中的 x 与 y 的实际资料求解的数值， a 和 b 的值确定了直线的位置，而它们一旦确定，这条直线就被唯一确定了。但用于描述这 n 组数据的直线有许多条，究竟用哪条直线来代表两个变量之间的关系，则需要有一个原

则。即希望这条直线离各离散点最近，具体表述为对于相关表中的 x ，它对应的实际数值 y 同这条直线上的理论值 y_c 的离差平方和为最小值。用公式表示为：

$$\sum (y - y_c)^2 = \sum (y - a - bx)^2 = \text{最小值}$$

设 $Q = \sum (y - a - bx)^2$ ，则 Q 是两个待定参数 a 和 b 的函数。要使 Q 为最小值，就要用对二元函数求极值的原理，求 Q 关于 a 和 b 的偏导数，并令其等于零。经过整理得出直线回归方程中求解参数 a 和 b 的标准方程组：

$$\begin{aligned} \sum y &= na + b \sum x \\ \sum xy &= a \sum x + b \sum x^2 \end{aligned}$$

解此方程组，得出求解 a 、 b 的计算公式为：

$$\begin{cases} b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} \\ a = \frac{\sum y}{n} - b \frac{\sum x}{n} \end{cases} \tag{9-5}$$

根据这一思想确定未知参数 a 和 b 的方法，称为最小平方法（Least squares analysis）。

【实例 9.3】 仍以表 9.1 给出的 2013 年我国部分省市人均 GDP 及人均消费水平资料为例，运用最小平方法，求解回归方程参数 a 和 b ，并建立一元线性回归模型。

解：设所配合的一元线性回归方程为： $y_c = a + bx$ ，其中，设人均 GDP 为 x ，人均消费水平为 y 。根据表 9.1 资料可得如表 9.4 所示的结果。

表 9.4 直线回归模型计算表（1） 单位：元

编 号	人均 GDP x	人均消费水平 y	x^2	xy
1	22 922	9 541	525 418 084	218 698 802
2	37 509	12 978	1 406 925 081	486 791 802
3	42 692	13 206	1 822 606 864	563 790 552
4	56 323	16 728	3 172 280 329	942 171 144
5	68 462	24 771	4 687 045 444	1 695 872 202
6	90 092	39 223	8 116 568 464	3 533 678 516
7	93 213	33 337	8 688 663 369	3 107 441 781
合 计	411 213	149 784	28 419 507 635	10 548 444 799

根据最小平方法确定参数的公式，可计算出参数 a 和 b 为：

$$\begin{cases} b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = 0.410 4 \\ a = \frac{\sum y}{n} - b \frac{\sum x}{n} = -2 711.12 \end{cases}$$

将 a 和 b 代入回归方程, 则有: $y_c = -2\,711.12 + 0.410\,4x$

回归直线确定后, 将各省市人均 GDP 数值依次代入方程式, 即可求得各省市相应的人均消费水平理论值。例如, 上海人均消费水平为 $-2\,711.12 + 0.410\,4 \times 90\,092 = 34\,262.64$ (元), 浙江人均消费水平为 $-2\,711.12 + 0.410\,4 \times 68\,462 = 25\,385.68$ (元)。

根据所建立的直线回归方程, 也可以进行外推预测。例如, 人均 GDP 为 100 000 元, 在其他条件相对稳定时, 可以预测其人均消费水平为: $y_c = -2\,711.12 + 0.410\,4 \times 100\,000 = 38\,328.88$ (元)。

在这里, 下列几个问题应引起注意:

(1) 回归方程中, $a = -2\,711.12$, 是人均消费水平的起点值, 即人均 GDP $x=0$ 时, y_c 的值。

(2) b 称为回归系数 (regression coefficient)。 $b=0.410\,4$ 则表示人均 GDP 每增加一个单位 (1 元), 人均消费水平平均增加 0.410 4 个单位 (元)。

(3) 由于 b 的计算公式中的分子与相关系数 r 计算公式中的分子完全一样, 且 b 与 r 为正值还是负值, 其符号均取决于这个分子, 所以, 回归系数 b 与相关系数 r 的符号必一致。即通过回归系数 b 的符号可判断两变量相关方向。

(4) 回归系数 b 与相关系数 r 之间还存在密切数量关系, 两者可相互推算。根据相关系数 $r = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y}$ 和回归系数 $b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = \frac{\sigma_{xy}^2}{\sigma_x^2}$, 可知 $\frac{b}{r} = \frac{\sigma_y}{\sigma_x}$, 亦即 $b = r \cdot \frac{\sigma_y}{\sigma_x}$ 。

(5) 另外, 一个直线回归方程只能做一种推算, 不能反向进行推算。也就是说, 只能以自变量 x 推算因变量 y , 而不能以因变量 y 推算自变量 x 。例如, 上例所配合的直线回归方程, 只能以给定的人均 GDP 数值时来推算人均消费水平, 不能以给定的人均消费水平数值反过来推算人均 GDP。

若根据研究需要, 想了解当人均消费水平一定时, 相对应的人均 GDP 数值应达到多少, 则应建立 x 对 y 的直线回归方程: $x_c = c + dy$ 。式中 c 与 d 的意义与公式 $y_c = a + bx$ 中 a 、 b 意义相同, 均为待定参数, 只是 x 与 y 的位置互换而已。同样采用最小平方法, 确定出参数 c 与 d 公式为:

$$\begin{cases} d = \frac{n \sum xy - \sum x \sum y}{n \sum y^2 - (\sum y)^2} \\ c = \frac{\sum x}{n} - d \frac{\sum y}{n} \end{cases} \quad (9-6)$$

【实例 9.4】 根据表 9.1 给出的 2013 年我国部分省市人均 GDP 及人均消费水平资料, 预测若某地人均消费水平为 20 000 元时, 相对应的人均 GDP 数值。

解: 若要预测人均消费水平为 20 000 元时, 相对应的人均 GDP 数值, 需建立以人均消费水平 y 为自变量, 人均 GDP x 为因变量的直线回归模型, 然后将自变量数值代入方程, 得到因变量的估计值, 如表 9.5 所示。

表 9.5 直线回归模型计算表（2）

单位：元

编 号	人均 GDP x	人均消费水平 y	y^2	xy
1	22 922	9 541	91 030 681	218 698 802
2	37 509	12 978	168 428 484	486 791 802
3	42 692	13 206	174 398 436	563 790 552
4	56 323	16 728	279 825 984	942 171 144
5	68 462	24 771	613 602 441	1 695 872 202
6	90 092	39 223	1 538 443 729	3 533 678 516
7	93 213	33 337	1 111 355 569	3 107 441 781
合计	411 213	149 784	3 977 085 324	10 548 444 799

根据最小平方方法确定参数的公式，可计算出参数 c 与 d 为：

$$\begin{cases} d = \frac{n \sum xy - \sum x \sum y}{n \sum y^2 - (\sum y)^2} = 2.2659 \\ c = \frac{\sum x}{n} - d \frac{\sum y}{n} = 10\,259.6 \end{cases}$$

将 c 和 d 代入回归方程，则有： $x_c = 10\,259.63 + 2.2659y$ 。显然，此时的回归方程与原来的回归方程是两条不同的回归直线，具有不同的斜率和意义，只能给定自变量来推算因变量。当人均消费水平为 20 000 元时，相对应的人均 GDP 理论数值为：

$$10\,259.63 + 2.2659 \times 20\,000 = 55\,577.63 \text{（元）}$$

9.2.4 估计标准误差

根据直线回归方程，按给定的自变量值可以推算出相应的因变量值，即得出估计值 y_c 。而估计值 y_c 与其对应的实际观察值 y 之间可能一致，也可能不一致，它们之间存在一系列离差，有的是正差，有的是负差，这些离差称为统计估计误差。这种估计误差的大小，可以说明推算结果的准确程度，即回归方程估计的准确程度。同时，统计上，一般是通过计算“估计标准误差”指标来反映回归方程的代表性的。若误差大，说明回归方程的代表性低；若误差小，则说明回归方程的代表性高。

估计标准误差（standard error of the estimate）是指因变量实际值与理论值离差的平均值，其计算原理与能够反映平均数代表性大小的标准差基本相同，定义公式为：

$$S_{yx} = \sqrt{\frac{\sum (y - y_c)^2}{n - 2}} \quad (9-7)$$

式中： S_{yx} ——估计标准误差，其下标 yx 代表 y 依 x 而回归的方程；

y_c ——根据回归方程推算出来的因变量的估计值；

y ——因变量的实际值；

n ——数据的项数。

利用定义式计算估计标准误差, 需要计算所有的估计值, 计算量大, 计算比较麻烦。将定义式中 y_c 用 $a+bx$ 替代, 经过化简, 可得如下的计算公式:

$$S_{yx} = \sqrt{\frac{\sum y^2 - a \sum y - b \sum xy}{n-2}} \quad (9-8)$$

【实例 9.5】 现仍以表 9.1 的有关资料为例, 计算估计标准误差。

解: 根据表 9.1 和表 9.5 中的数据资料可得如下计算结果:

$$\sum y^2 = 3\,977\,085\,324, \quad \sum y = 149\,784, \quad \sum xy = 10\,548\,444\,799;$$

根据前面计算有: $a = -2\,711.12$, $b = 0.410\,4$; 由 $n=7$, 有

$$S_{yx} = \sqrt{\frac{\sum y^2 - a \sum y - b \sum xy}{n-2}} = 3\,288.96$$

结果表明, 7 个省市人均消费水平的估计理论值与实际值的平均误差为 3 288.96 元。由此可见, 只有把回归估计值与估计标准误差结合起来分析运用, 才更具有意义。

估计标准误差 S_{yx} 与相关系数 r 在数量上也存在着密切的关系, 可看成是从另一个角度说明相关分析与回归分析之间的联系。两者之间的关系可由下列公式表述:

$$r = \sqrt{1 - \frac{S_{yx}^2}{\sigma_y^2}} \quad (9-9)$$

$$S_{yx} = \sigma_y \sqrt{1 - r^2} \quad (9-10)$$

从相互联系的两个算式中可以看出, r 与 S_{yx} 的变化方向是相反的。当 r 越大时, S_{yx} 越小, 这时相关密切程度较高, 回归直线的代表性较大; 当 r 越小时, S_{yx} 越大, 这时相关密切程度较低, 回归直线的代表性较小; $r = \pm 1$ 时, $S_{yx} = 0$, 这时现象间完全相关, 各相关点均落在回归直线上, 此时对 x 的任何变化, y 总有一个相应的值与之对应; $r=0$ 时, S_{yx} 取得最大值, 这时现象间不存在直线关系。

9.2.5 判定系数

在回归分析中, 除了可用估计标准误差反映回归方程估计的准确程度及回归直线代表性的大小外, 还有一个广泛应用的指标, 叫作判定系数 (coefficient of determination)。它通常用 r^2 表示, 用来测定回归方程拟合数据的好坏程度。判定系数 r^2 就是相关系数 r 的平方。当然, 判定系数有它的基本公式: $r^2 = \frac{\sum (y_c - \bar{y})^2}{\sum (y - \bar{y})^2}$ 。式中, 分母 $\sum (y - \bar{y})^2$ 是总离差平方和, 分子

$\sum (y_c - \bar{y})^2$ 叫回归平方和。 r^2 越大, 回归平方和在总离差平方和中所占的比重就越大, 表明总离差中由回归方程来解释的部分也越大, 线性回归效果就越好。

判定系数的范围在 0 与 1 之间。如果 r^2 为 1，则表明两个变量之间有非常好的相关性， y 的估计值与实际值之间没有差别，此时，估计标准误差为 0；如果 r^2 为 0，则回归方程不能用来预测 y 值，此时，估计标准误差最大。

例如，根据表 9.1 给出的 2013 年我国部分省市人均 GDP 及人均消费水平资料，前面已算出 $r=0.964\ 3$ ，则判定系数 $r^2=0.929\ 9$ 。这说明，总离差平方和中 92.99% 可以用回归方程来解释，人均 GDP 与人均消费水平两者之间高度相关，回归直线拟合的数据较好。

9.3 应用相关分析和回归分析应注意的问题

9.3.1 在定性分析的基础上进行定量分析

相关关系的有无、自变量和因变量的确定、相关的类型及计算结果的意义等问题，都必须根据人们对研究对象的了解，做出定性判断。因为相关分析的方法，若不能解释相关关系产生的原因，它本身不能判断现象之间是否存在质的关系，只有依靠定性分析并在此基础上进行定量分析，才能对现象之间的关系做出科学的判断。否则，若对本来没有内在关系的现象进行相关分析，将会导致虚假相关的错误。例如，研究某地区人口出生数量与该地区老年人口再婚数量之间的关系，显然是没有任何意义的。若据此继续进行回归分析并对现象进行推算预测，其结论将是荒谬的。

9.3.2 要注意现象质的界限及相关关系作用的范围

在进行相关分析与回归分析时，要注意现象质的界限及相关关系作用的范围。超出了这个范围，分析的结果就会与事实不符。特别是利用回归方程进行预测时，不能无限外推。因为在预测中，指定 x 变量值未超出所用资料的数值范围，叫作内插回归预测，其预测效果好；在预测中，指定 x 变量值超出所用资料的数值范围，叫作外推回归预测，预测效果随着 x 变量值对所用资料的数值范围的远离而逐渐降低。回归分析中，回归直线方程是采用最小平方方法取得的，最小平方方法指的是对现有资料范围配合一条“最佳”线，如果外推到范围以外，则不一定是最佳线了。根据样本数据所建立的回归方程，只有代表经济变量之间的数量条件才能够成立。如果忽视相关关系建立的条件，把这种关系无限制地向外推广，所得结论是不科学的。例如，商品的价格与商品的销售量只有在一定的范围内才具有负相关关系，商品的价格超过一定限度，销售量有可能与之作同方向变化。社会经济现象中，许多现象也是如此。因此，用相关分析与回归分析方法进行推算和预测时，要注意它的作用范围。

9.3.3 要将各种分析指标结合应用

相关分析与回归分析中的各种分析指标，既有区别又有联系。相关分析中的相关系数，是反映现象之间直线相关关系密切程度的指标；回归分析中的直线回归方程，反映具有相关关系两变量之间数量变化关系；而回归直线的代表性大小则由回归分析中的估计标准误差来反映。它

们分别从不同的层次和侧面揭示现象之间数量关系的特征。所以，在进行相关分析和回归分析时，只有将它们有机地结合起来，才能更加全面、准确、科学地描述社会经济现象之间数量变化的规律性。

9.3.4 要尽可能使用大样本材料

大量观测是统计分析的基础，只有使用大样本资料进行大量观测才能淡化偶然因素的影响，获得现象间的真实关系。不管是相关分析中通过计算相关系数判定相关密切程度，还是回归分析中利用回归方程进行统计预测，如果样本材料过少，都不易真实反映客观现象之间的关系。本章中给出的例题样本均较小，这只是为了写作上的方便，在实际中不宜采用。

9.4 Excel在相关回归分析中的应用

针对相关分析和回归分析，Excel 提供了更加方便、快捷的工具和统计函数。本节主要说明利用工具与统计函数进行相关分析与回归分析。

9.4.1 利用Excel进行相关分析

相关系数是测定相关程度的最重要的指标。在 Excel 中，测定相关系数有两种方法：相关系数函数法和相关系数工具。

【实例 9.6】 根据 1991—2005 年国内生产总值与国家内债发行情况的资料（单位：亿元），分别采用相关系数函数与相关系数工具进行二者之间的相关分析。

1) 用函数求相关系数

在 Excel 中，计算两个变量之间相关系数的函数为 CORREL 函数。采用此函数计算相关系数的步骤如下，数据资料及计算结果如图 9.7 所示。



The screenshot shows an Excel spreadsheet with the following data:

1	A	B	C	D	E	F	G	H
2	年 份	国内生产总值	国内债务					
3	1991	21781.5	199.3					
4	1992	26923.5	395.64					
5	1993	35333.9	314.78					
6	1994	48197.9	1028.57					
7	1995	60793.7	1510.86					
8	1996	71176.6	1847.77					
9	1997	78973.0	2412.03					
10	1998	84402.3	3228.77					
11	1999	89677.1	3702.13					
12	2000	99214.6	4153.59					
13	2001	109655.2	4483.53					
14	2002	120332.7	5660					
15	2003	135822.8	6029.24					
16	2004	159878.3	6726.28					
17	2005	183084.8	6922.87					

In cell E5, the formula `=CORREL(B2:B17,C2:C17)` is entered, resulting in the value 0.9802689 in cell F5.

图 9.7 利用 CORREL 函数计算相关系数

第一步：单击任一空白单元格，本例中选定的单元格为 F4，单击“插入”菜单，选择“函数”选项，弹出“粘贴函数”对话框。在对话框的左侧“函数分类”中选择“统计”，在右侧的“函数名”中选择 CORREL。单击“确定”按钮，出现 CORREL 对话框，如图 9.8 所示。

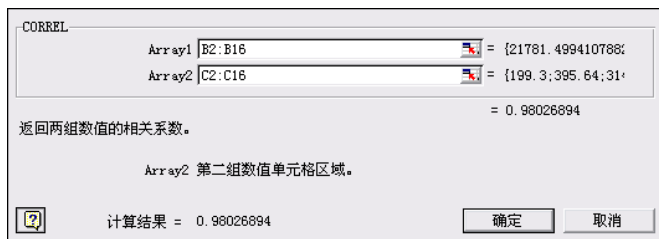


图 9.8 CORREL 对话框

第二步：在 CORREL 对话框中的“Array1”和“Array2”框中分别输入用于计算相关系数的两组数据所在的单元格区域。本例中，在“Array1”中输入“B2:B16”，在“Array2”中输入“C2:C16”，即可在对话框下方显示出计算结果为 0.98 026 894。

第三步：单击“确定”按钮，即可在选定的空白单元格显示出相关系数。本例在 F4 单元格显示相关系数 0.9 802 689。

2) 相关系数工具

采用相关系数工具计算相关系数的操作步骤如下。

第一步：单击“工具”菜单，选择“数据分析”选项。打开“数据分析”对话框，从其对话框的“分析工具”列表中选择“相关系数”，单击“确定”按钮，打开“相关系数”对话框，如图 9.9 所示。



图 9.9 “相关系数”对话框

第二步：在“相关系数”对话框中确定输入区域和输出选项。在“输入区域”框中输入分析数据所在的单元格区域。在本例中，输入区域为\$B\$1:\$C\$16。“分组方式”中指出输入区域中的数据是按行还是按列排列，本例为“逐列”。若输入区域包括列标志行，则选中“标志位于第一行”复选框，本例选中此复选框。

在“输出选项”对话框中可以指定结果的输出去向，输出去向有三种。在“输出区域”框中输入输出结果所在的单元格区域，在本例中，输出区域为\$F\$4。也可以通过选择“新工作表组”或“新工作簿”将结果放在新工作表或新工作簿。

第三步：单击“确定”按钮，在指定位置给出计算结果，如图9.10所示。

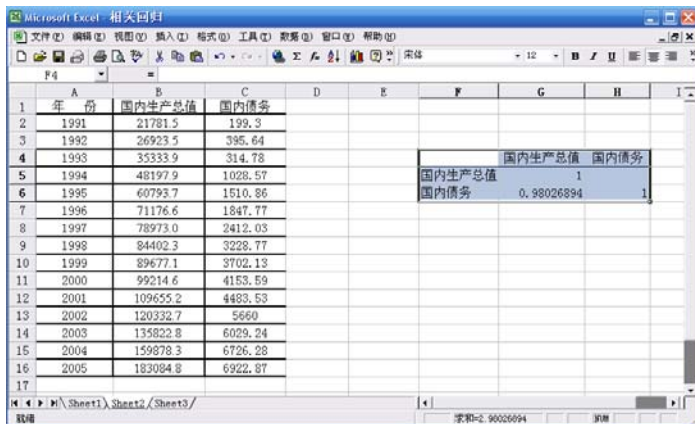


图 9.10 利用相关系数工具计算相关系数结果

在图9.10的输出结果中，国内生产总值和国内债务的自相关系数均为1，国内生产总值和国内债务的相关系数为0.9803，与用函数计算的结果完全相同。

9.4.2 利用Excel进行回归分析

通过相关分析，可以观察到变量之间存在相关关系及紧密程度，如果想进一步确定它们之间的数量变化关系，就要借助于回归分析。以下通过一个线性回归实例来说明在Excel中如何利用回归分析工具进行回归分析的操作方法。

【实例9.7】 根据1991—2005年国内生产总值与国家内债发行情况的资料进行回归分析。

采用回归分析工具进行回归分析的操作步骤如下。

第一步：单击“工具”菜单，选择“数据分析”选项。打开“数据分析”对话框，从其对话框的“分析工具”列表中选择“回归”选项，单击“确定”按钮，打开“回归”对话框，如图9.11所示。

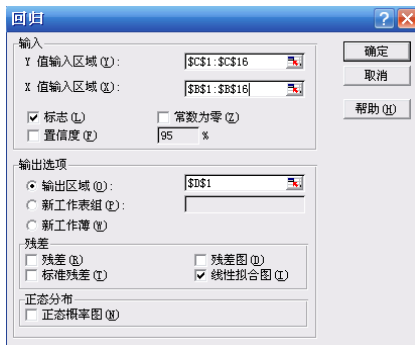


图 9.11 “回归”对话框

第二步：在“回归”对话框中确定输入区域和输出选项。

确定输入区域。在“Y 值输入区域”框中输入因变量数据所在的单元格区域，在本例中，输入区域为\$C\$1:\$C\$16。在“X 值输入区域”框中输入自变量数据所在的单元格区域，在本例中，输入区域为\$B\$1:\$B\$16。

输入选项中其他复选框说明：如果有标志，则选中“标志”复选框；若要求回归直线通过原点，则选中“常数为 0”复选框；若要求输出置信度，则选中“置信度”复选框。

确定输出选项。在“输出选项”对话框中可以指定结果的输出去向，输出去向有三种。在“输出区域”框中输入输出结果所在的单元格区域的左上角的单元格的行列号，在本例中，输出区域为\$D\$1。也可以通过选择“新工作表组”或“新工作簿”将结果放在新工作表或新工作簿。

输出选项中其他复选框说明：若要求输出残差、标准残差、残差图、线性拟合图和正态概率图，则选中相应的复选框。本例中选中“线性拟合图”。

第三步：单击“确定”按钮后，在指定位置给出计算结果，如图 9.12 所示。

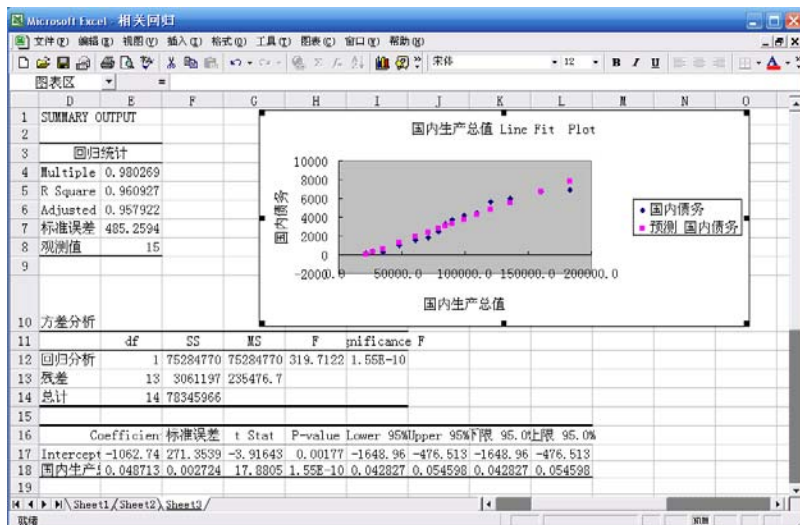


图 9.12 回归分析结果



统计术语

相关 correlation

相关表 correlation table

散点图 scatter plot

回归系数 regression coefficient

回归直线 regression line

最小平方方法 least squares analysis

简单线性回归 simple linear regression

因变量 independent variable

回归 regression

相关图 correlation diagram

相关系数 correlation coefficient

判定系数 coefficient of determination

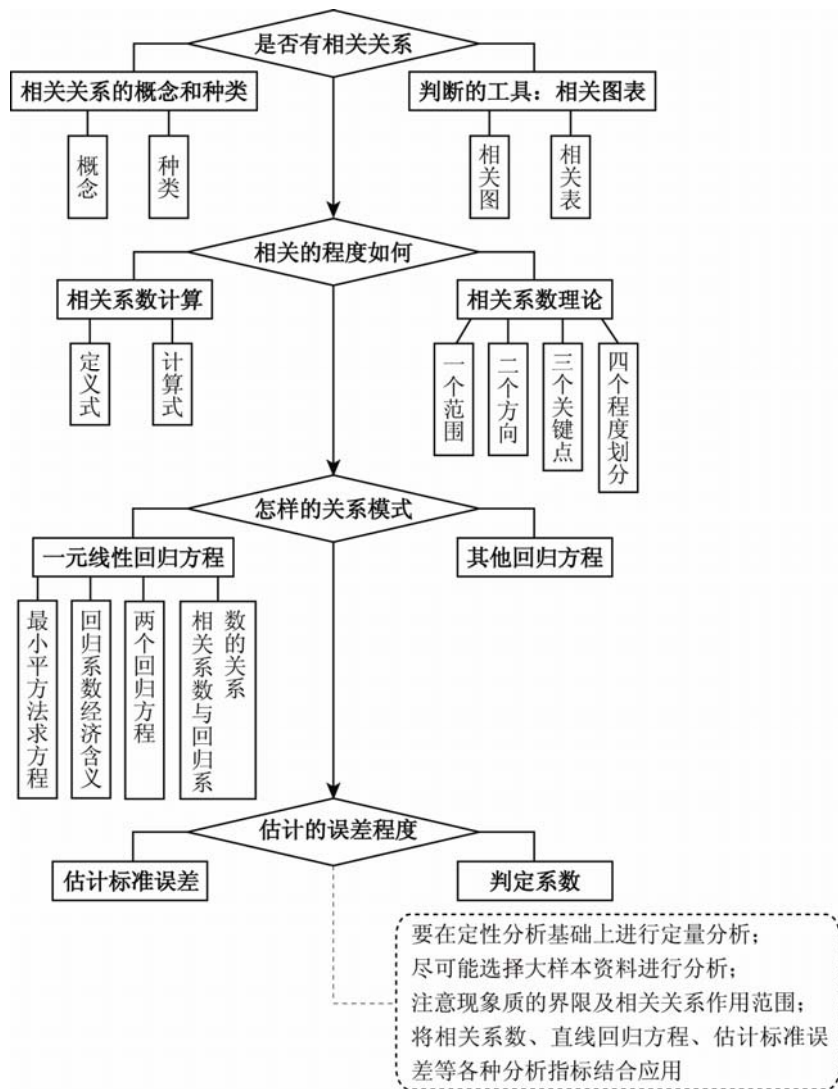
回归方程 regression equation

估计标准误差 standard error of the estimate

自变量 dependent variable



重点知识梳理



习题与实践训练

一、判断题

1. 正相关指的是两个相关变量的数量变动方向都是上升的。 ()
2. 相关系数越小，两变量之间相关的密切程度越低。 ()
3. 若变量 x 的值减少时，变量 y 的值也减少，说明变量 x 与 y 之间存在正的相关关系。 ()
4. 若直线回归方程为 $y_c = 80 - 9.6x$ ，则变量 x 与 y 之间存在负的相关关系。 ()
5. 由变量 y 依变量 x 回归和由变量 x 依变量 y 回归所得到的回归方程之所以不同，主要是因为方程中参数表示的意义不同。 ()

6. (甲) 某产品产量与单位成本的相关系数是 -0.8988 ; (乙) 产品单位成本与利润率的相关系数是 -0.9567 ; 因此, (乙) 比 (甲) 的相关程度高。 ()

7. 当相关系数 r 的绝对值等于 0 时, 只能说明 x 与 y 两个变量之间无线性相关, 不能说明两个变量不相关。 ()

8. 只有当相关系数接近+1 时, 才能说明两变量之间存在高度相关关系。 ()

9. 回归系数 b 与相关系数 r 不仅符号完全一致, 而且存在数量关系。 ()

10. 判定系数越大, 表明线性回归效果就越好。 ()

二、单项选择题

1. 如果相关系数 $r=-0.8566$, 说明两个变量之间 ()。

- A. 完全相关 B. 完全不相关 C. 高度相关 D. 显著相关

2. 当所有的观察值 y 都落在直线 $y_c=a+bx$ 上时, 则 x 与 y 之间的相关系数为 ()。

- A. $-1 < r < 1$ B. $0 < r < 1$ C. $r=1$ D. $r=0$

3. 变量之间的相关程度越低, 则相关系数的数值 ()。

- A. 越接近于 -1 B. 越接近于 $+1$ C. 越接近于 0 D. 越小

4. 学院对学生考试成绩与学习时间进行测定分析, 建立了考试成绩与学习时间的直线回归方程: $y_c=160-2.3x$, 该方程明显有错, 错误在于 ()。

- A. 自变量和因变量的关系搞错了
B. a 值的计算是正确的, b 值是错误的
C. b 值的计算是正确的, a 值是错误的
D. a 值和 b 值的计算都有错误

5. 单位成本 (元) 与产量 (千件) 变动的回归方程为: $y_c=77.37-1.82x$, 这意味着 ()。

- A. 产量每增加 1 千件, 单位成本增加 75.52 元
B. 产量每增加 1 千件, 单位成本减少 1.82 千元
C. 产量每增加 1 千件, 单位成本减少 1.82 元
D. 如果产量增加 1 千件, 则单位成本为 77.37 元

6. 估计标准误差说明回归直线的代表性, 因此 ()。

- A. 估计标准误差越大, 说明回归直线的代表性越大
B. 估计标准误差越小, 说明回归直线的代表性越小
C. 估计标准误差越大, 说明回归直线的代表性越小
D. 估计标准误差越小, 说明回归直线的实用价值小

7. 若已知某企业有如下资料:

产量 (千件)	2	3	4	3	4	5
单位产品成本 (元/件)	73	72	71	73	69	68

则不需要计算, 便知道该企业产量与单位产品成本的相关系数为 ()。

- A. 0.9087 B. -0.9091 C. 1.0323 D. -0.4998

8. 相关系数的取值范围是 ()。

- A. $0 \leq r \leq 1$ B. $-1 \leq r \leq +1$ C. $-1 \leq r \leq 0$ D. $-1 < r < 1$

9. 如果判定系数为 1, 则表明两个变量之间 ()。
- A. 估计标准误差很大, 回归直线拟合的数据很差
 - B. 相关系数绝对值必为 1, 估计标准误差必为 0
 - C. 有高度相关关系, 估计标准误差必为 0
 - D. 估计标准误差较大, 回归直线拟合的数据很好
10. 下面哪两个变量之间的相关程度高 ()。
- A. 商品的广告支出费用与商品的销售收入的相关系数是 0.89
 - B. 汽车的维护费用与行驶里程的相关系数是 0.902 1
 - C. 警务支出费用与犯罪率的相关系数是 -0.876 7
 - D. 平均流通费用率与商业利润率的相关系数是 -0.94

三、多项选择题

1. 请指出下列哪些关系是相关关系 ()。
- A. 家庭收入增多, 其消费支出也有增长的趋势
 - B. 农作物的收获量和雨量、气温、施肥量有着密切的关系
 - C. 物价上涨, 商品的需求量下降
 - D. 一个地区经济越发达, 妇女的生育率越低
 - E. 圆的直径越长, 其面积和周长都越大
2. 直线回归方程中, 两个变量 x 和 y ()。
- A. x 是自变量, y 是因变量
 - B. x 是给定的变量, y 是随机变量
 - C. 两个都是给定的变量
 - D. 两个都是随机变量
 - E. 两个变量完全不相关
3. 员工的薪酬(元) y 与劳动生产率(千元) x 的回归方程为: $y_c = 60 + 100x$, 这意味着 ()。
- A. 如果劳动生产率等于 1 000 元, 则员工薪酬为 100 元
 - B. 劳动生产率每增加 1 000 元, 则员工薪酬平均提高 100 元
 - C. 劳动生产率每增加 1 000 元, 则员工薪酬增加 160 元
 - D. 如果劳动生产率等于 1 000 元, 则员工薪酬为 160 元
 - E. 劳动生产率每下降 1 000 元, 则员工薪酬平均减少 100 元
4. 直线回归方程 $y_c = a + bx$ 中的 b 称为回归系数, 回归系数的作用是 ()。
- A. 可确定两变量之间的相关程度
 - B. 可确定两变量之间的相关方向
 - C. 可确定因变量的实际值与估计值的变异程度
 - D. 可确定当自变量增加一个单位时, 因变量的平均增加值
 - E. 可确定当自变量为 0 时, 因变量的起点值
5. 在直线回归方程中 ()。
- A. 一个回归方程只能做一种推算
 - B. 两个变量中须确定自变量和因变量

- C. 回归系数只能取正值
- D. 回归系数与相关系数的符号是一致的
- E. 同一方程中, 给定 x 的值可以推测 y , 给定 y 的值可以反过来推测 x
- 6. 进行相关分析和回归分析时要注意的问题有 ()。
 - A. 先进行定量分析, 然后进行定性分析
 - B. 要注意现象质的界限及相关关系作用的范围
 - C. 将相关系数、直线回归方程、估计标准误差等各种分析指标结合应用
 - D. 为计算方便, 尽量选择小样本资料
 - E. 利用回归方程进行预测时, 不能无限制地向外推测
- 7. 下列哪种情况下, 回归直线代表性较好 ()。
 - A. 估计标准误差为 0
 - B. 相关系数的绝对值接近于 1
 - C. 判定系数为 0
 - D. 判定系数为 1
 - E. 相关系数为 0
- 8. 估计标准误差、判定系数及相关系数三者的关系是 ()。
 - A. 估计标准误差越大, 判定系数越大
 - B. 估计标准误差越大, 判定系数越小
 - C. 判定系数为 0 时, 估计标准误差最大
 - D. 判定系数为 1 时, 相关系数最大
 - E. 相关系数为 1 时, 判定系数最大
- 9. 在回归分析中, 就两个相关变量 x 和 y 而言, 由变量 y 依变量 x 回归和由变量 x 依变量 y 回归所得到的回归方程是不同的, 这种不同表现在 ()。
 - A. 方程中参数的数值不同
 - B. 方程中参数估计的方法不同
 - C. 估计标准误差的计算方法不同
 - D. 估计标准误差的数值不同
 - E. 参数表示的实际意义不同
- 10. 可以用来判断现象之间相关方向的指标有 ()。
 - A. 估计标准误差
 - B. 判定系数
 - C. 相关系数
 - D. 回归系数
 - E. 两变量的协方差

四、填空题

1. 完全相关即为_____关系, 其相关系数为_____。
2. 当变量 x 按一定数额变动时, 变量 y 也按一定的数额变动, 这时变量 x 与 y 之间存在着_____关系。
3. 用来说明回归方程代表性大小的统计分析指标有_____。
4. 直线回归方程 $y_c = a + bx$ 中的参数 a 表示的意思是_____, b 是_____, 其含义是_____, 估计特定参数使用的方法是_____法。
5. 相关分析是回归分析的_____, 而回归分析是相关分析的_____。
6. 标准差是说明_____的代表性的, 而估计标准误差是反映_____的代表性大小的。

- 7. 相关系数 $r=-1$ ，说明两变量之间_____； $r=+1$ ，说明两变量之间_____； $r=0$ ，说明两变量之间_____。
- 8. 判定系数的范围在_____之间。
- 9. 判定系数为 0，则回归直线_____；判定系数为 1，则回归直线_____。
- 10. 相关系数的四个相关密切程度等级划分是_____、_____、_____和_____。

五、应用能力训练题

1. 随机抽取 10 家航空公司，对其最近一年的航班正点率和顾客投诉次数进行了调查，所得数据如下。

航班正点率 (%)	81.8	76.6	76.6	75.7	73.8	72.2	71.2	70.8	91.4	68.5
顾客投诉次数	21	58	85	68	74	93	72	122	18	125

- (1) 编制相关表，绘制相关图，说明航班正点率和顾客投诉次数之间的关系。
 - (2) 计算相关系数。
 - (3) 用航班正点率作为自变量，顾客投诉次数作为因变量，求出估计的回归方程，并解释回归系数的意义。
 - (4) 如果航班正点率为 80%，估计顾客的投诉次数。
2. 某年某地区有 10 个品牌啤酒的广告费用和销售量的数据如下。

广告费用 (万元)	120	68.7	100.1	76.6	8.7	1	21.5	1.4	5.3	1.7
销售量 (万箱)	36.3	20.7	15.9	13.2	8.1	7.1	5.6	4.4	4.4	4.3

根据上面材料：

- (1) 用广告费支出作自变量，销售量作因变量，求出估计的回归方程。
 - (2) 广告费用每增加 10 000 元，销售量将有何变化？
3. 航空公司经过统计调查得知，波音 737 飞机 500 英里飞行成本与乘客数量之间存在相关关系，调查数据资料如下。

乘客数 (人)	61	63	67	69	70	74	76	81	86	91	95	97
成本 (千美元)	4.28	4.08	4.42	4.17	4.48	4.30	4.82	4.70	5.11	5.13	5.64	5.56

- (1) 建立飞行成本 y 与乘客数量 x 的直线回归方程。
 - (2) 飞行每增加一名乘客，飞行成本将增加或减少多少美元？
 - (3) 在此案例中，如果飞行时不搭载乘客，飞行 500 英里的飞行成本为多少美元？
4. 在某些国家和地区，在饭店就餐有给服务员小费的习惯。几乎所有人都相信，在账单的数额和小费之间存在着联系：较大的账单就会有较多的小费。现搜集到 6 次午餐会的成对数据如下。

账单 (美元)	33.46	50.68	87.92	98.84	63.60	107.34
小费 (美元)	5.50	5.00	8.08	17.00	12.00	16.00

- (1) 是否有足够的证据判断：在账单数额和小费数额之间存在某种联系？（它们相关的程度如何？）

(2) 若存在某种联系，则使用这种联系来确定当账单数额为 100 美元时，应该留下多少小费？

(3) 现已知给出小费 10 元，请推测其消费的账单数额为多少？

5. 某种饮料有如下销售资料：

时期	1	2	3	4	5	6	7	8	9	10
销售量（箱）	430	335	520	490	470	210	195	270	400	480
气温（℃）	30	21	35	42	37	20	8	17	35	25

(1) 销售量与气温之间是否具有关系？怎样判断？

(2) 若具有关系，其关系的密切程度如何？

(3) 用最小二乘法配合销售量 y 随气温 x 变动的直线回归方程。

(4) 气温每升高 1℃，销售量平均增加（减少）多少箱？

6. GPA 英语全称是 grade point average，意思就是平均绩点（平均分）。下表是 10 位优秀学生的学习时间和平均学分资料，请计算相关系数，并思考相关系数为什么这么低？

学习时间	23	12	15	14	16	21	14	11	18	9
GPA	3.95	3.90	4.00	3.76	3.97	3.89	3.66	3.91	3.80	3.89

7. 将学生分为 8~12 人的小组：

(1) 记录下每个小组成员的身高，再测量出他们的体重并记录下来，在身高和体重之间存在一种关系吗？如果存在，这种关系是什么？

(2) 测量每个小组成员的身高和手臂伸开的宽度。测量臂宽时，测试对象应站直，将手臂平伸，如同飞机上的机翼。在身高与臂宽之间是否存在相关性？如果存在，求出臂宽与身高的回归方程。

(3) 记录下每个小组成员在一分钟内心跳的次数、心率，再记录下他（她）的身高。在心率和身高之间存在一种关系吗？如果存在，这种关系是什么？

8. 在第 2 章统计调查应用能力训练题中，我们搜集到了你所在班级同学的身高、体重原始资料，经过统计整理后，形成条理的统计资料。请根据这些材料，利用本章学过的知识，判断你班同学身高和体重之间相关程度如何？是否近似直线相关？是哪条直线？

9. 某公司所属 8 个企业的产品销售资料如下表所示。

企业编号	产品销售额（万元）	销售利润（万元）
1	170	8.1
2	220	12.5
3	390	18.0
4	430	22.0
5	480	26.5
6	650	40.0
7	950	64.0
8	1 000	69.0

要求分别利用 Excel 软件完成以下统计分析：

- (1) 计算产品销售额与利润额之间的相关系数。
- (2) 确定利润额对产品销售额的直线回归方程。

10. 某学校大一女生的体重与肺活量的数据资料如下。

体重 (Kg)	42	42	46	46	46	50	50	50	52	52	58	58
肺活量 (L)	2.55	2.20	2.75	2.40	2.80	2.81	3.41	3.10	3.46	2.85	3.80	3.00

要求：

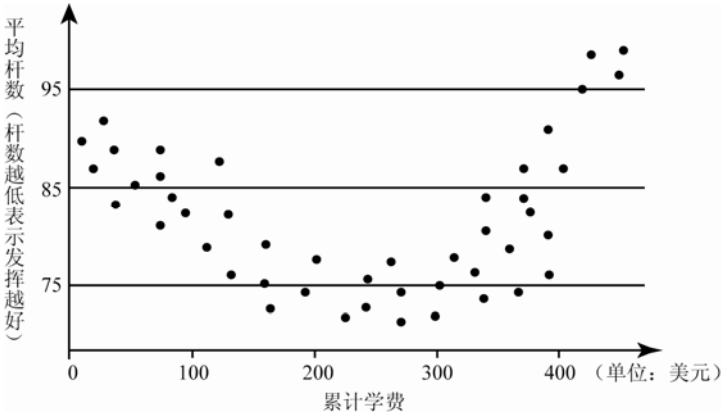
- (1) 利用 Excel 软件计算体重与肺活量之间的相关系数。
- (2) 利用 Excel 软件配合体重与肺活量的直线回归方程。



本章案例

回归分析可以说是统计学弹药库中的“氢弹”。无论是谁，只要有一台电脑和一个大型样本数据，在家中或办公室里就能成为一个研究员。这样做会出现什么错误？各种错误。回归分析为复杂的问题提供了精确的答案，但这些答案却不一定准确。在错误运用这一统计工具的人的手中，回归分析会得出误导甚至错误的结果；但即使在正确运用这一统计工具的人的手中，这一强大的统计工具依然可以让我们在错误的方向上危险超速、越走越远。下面就是让回归分析这一非凡工具沦为“邪恶”工具的常见错误。

用回归方程式来分析非线性关系。不知道你是否留意过吹风机上贴的那句警示标语——禁止在浴缸中使用。如果你看到了，心里一定会想：“有谁会傻到在浴缸里吹头发呢？”吹风机是电器，我们都知道在有水的环境中应该远离电器，因为会触电。假如回归分析也有一个类似的警示标语，那么上面写的就会是：当分析的两个变量之间不存在线性关系时禁止使用。请记住，回归系数所描述的是那条“数据最佳拟合直线”的坡度，一定要是一条直线，假如不是，就需要在另外的地方用不一样的坡度方程来诠释。举个例子，如下是我每个月上的高尔夫球课（解释变量）和每场 18 洞练习的平均杆数（因变量）之间的关系，如下图所示，两者之间并不存在一致的线性关系。



上图中并非完全没有规律，只不过是难以用一条直线来描述罢了。前几节高尔夫球课使我的杆数快速降了下来，因此，在这个阶段，我的课程数与杆数是呈负相关关系的，斜度为负。也就是说，上课降低了我的杆数（对于打高尔夫球来说这是一件好事）。

但是，当我的学费累计交到了 200 ~ 300 美元时，这个阶段的课程似乎对我的球场表现没有太大的帮助。高尔夫球课程与我的成绩之间似乎不存在一个明确的关系，因此斜度为零。

随着上课的次数越来越多，我的成绩甚至出现了下滑。当累计学费达到了 300 美元以上，增加的课程反而使我的杆数越来越高，在这个阶段斜度就为正了。最重要的一点是，我们无法用一个系数来准确概括高尔夫球课程和成绩之间的关系。对于上述关系来说，一个最佳的描述方式是：高尔夫球课程与我的挥球杆数之间存在着若干个不同的线性关系。你看到这种情况，但是电脑上的统计软件却看不到。如果你一股脑地把这些数据输入回归方程中，电脑也会生成一个系数，但这个系数将无法准确地反映不同变量之间的真正关系，这其实与在浴室里用吹风机是一样的。

只有当变量之间的关系为线性时，回归分析才可派上用场。无论什么统计工具，离它的初始功能偏差越大，其效果就会越差，有时候甚至还会有危险。

（资料来源：《赤裸裸的统计学》【美】Charles Wheelan 著，曹槟译，中信出版社，2013）

附录 A 【习题与实践训练】答案

第 1 章

一、1. × 2. × 3. × 4. × 5. × 6. × 7. √ 8. × 9. × 10. √

二、1. B 2. D 3. D 4. D 5. A 6. C 7. D 8. B 9. A 10. C

三、1. BCDE 2. ABCD 3. ACD 4. ADE 5. ABC 6. ABCDE 7. ADE 8. BCE

9. ABCDE 10. BCD

四、1. 工作过程与结果 理论与实践

2. 统计设计 统计调查 统计整理 统计分析

3. 大量观察法 统计分组法 综合指标法

4. 变动的数量标志

5. 总体单位 数量标志 品质标志 统计总体 数量指标 质量指标

6. 指标条件 指标名称 指标数值 指标单位

7. 一系列相互联系的统计指标所组成的有机整体 基本 专题

8. 定性数据 定量数据

9. 母体 子样

10. 连续型变量 离散型变量

五、1. 略。2. 略。

3. 轿车生产总量、旅游收入是数量指标；经济发展速度、人口出生率、安置再就业人数、城镇居民人均可支配收入、恩格尔系数是质量指标。

区分数量指标与质量指标的方法是：数量指标用绝对数表示，质量指标用相对数和平均数表示。

4. 总体：我国总人口；样本：1 400 万人；变量：年龄、住居面积数、生育子女数等。

5. 略。

6. 略。

7. 略。

8. 略。

第 2 章

一、1. × 2. × 3. × 4. √ 5. × 6. × 7. × 8. √ 9. × 10. ×

二、1. C 2. A 3. A 4. C 5. B 6. D 7. A 8. C 9. B 10. D

三、1. BDE 2. ACD 3. BCD 4. ADE 5. BCE 6. ABCE 7. ABCE 8. BDE

9. BDE 10. ABC

四、1. 单一表 一览表

2. 普查 抽样调查 重点调查 典型调查

3. 全面调查 非全面调查

4. 直接观察法

5. 有意识 随机

6. 统计调查方案
7. 表头 表体 表脚
8. 访问法（采访法、询问法）
9. 开放式 封闭式
10. 网上直接调查 网上间接调查

五、1. 略。2. 略。

3. (1) 全面调查 一次性调查 直接调查 统计报表
- (2) 全面调查 一次性调查 直接调查 普查
- (3) 非全面调查 经常性调查 直接调查 抽样调查
- (4) 非全面调查 经常性调查 直接调查 重点调查
- (5) 非全面调查 经常性调查 直接调查 典型调查
4. (1) 直接观察法 直接调查

- (2) 实验法 直接调查
 - (3) 访问法 直接调查
 - (4) 网上调查 直接调查
5. 略。6. 略。7. 略。

第3章

一、1. × 2. × 3. × 4. × 5. √ 6. √ 7. × 8. × 9. √ 10. √

二、1. B 2. B 3. C 4. A 5. D 6. A 7. D 8. B 9. D 10. A

三、1. ACD 2. BCD 3. ABDE 4. CD 5. ACE 6. BCE 7. BCD 8. ABE

9. BCDE 10. BC

- 四、1. 选择分组标志和划分各组界限
2. 类型分组 结构分组 分析分组
3. 品质分配数列 变量分配数列
4. 组限 每个组的最大值 每个组的最小值 组中值
5. 反比
6. 上组限不在内
7. 等距分组 异距分组
8. 按某标志分的组 各组相应的分配次数或频率或标志值 频数 频率
9. 钟型分布 U型分布 J型分布
10. 总标题 横行标题 纵栏标题 数字资料

五、1. 组距数列分布表

分 组	频 数	频率 (%)
200~300	2	6.25
300~400	13	40.625
400~500	12	37.5
500~600	3	9.375
600~700	2	6.25
合计	32	100.00

次数分布图略。

2. (1) 组距数列分布表

分 组	频 数	频率 (%)
650~660	2	2
660~670	5	5
670~680	6	6
680~690	14	14
690~700	26	26
700~710	18	18
710~720	13	13
720~730	10	10
730~740	3	3
740~750	3	3
合 计	100	100

(2) 直方图和次数分配曲线略。

3. 组距数列频率及累计频数分布表

分 组	频 率	频率 (%)	向上累积		向下累积	
			频数	频率 (%)	频数	频率 (%)
60 分以下	5	10	5	10	50	100
60~70	8	16	13	26	45	90
70~80	14	28	27	54	37	74
80~90	16	32	43	86	23	46
90~100	7	14	50	100	7	14
合 计	50	100	—	—	—	—

成绩呈钟型分布。60 分以下 5 个人，80 分以上 23 个人。

4. (1) 人力资源部门统计申报五个不同岗位的人数；(饼图、条形图)

(2) 某学院男女学生的比重；(饼图)

(3) 2005—2015 年公司每年主营业务收入数值；(趋势图)

(4) 八个项目中每个项目的得分。(条形图)

5. 略。

6. 略。

7. 频率与累积频率分布：

分 组	频数 f	频率 (%)	向上累积		向下累积	
			频数	频率 (%)	频数	频率 (%)
100~110	6	12	6	12	50	100
110~120	13	26	19	38	44	88
120~130	18	36	37	74	31	62
130~140	9	18	46	92	13	26
140~150	4	8	50	100	4	8
合 计	50	100	—	—	—	—

图略。

8. 略。

第4章

一、1. √ 2. × 3. × 4. × 5. × 6. × 7. × 8. × 9. √ 10. ×

二、1. B 2. A 3. D 4. B 5. B 6. D 7. A 8. D 9. C 10. C 11. D 12. C 13. A

14. B 15. D 16. B 17. D 18. B 19. A 20. D

三、1. ACE 2. ABC 3. ABCD 4. CDE 5. AC 6. BD 7. ABC 8. ABC 9. ABE 10. BE

四、1. 总体单位总量 总体标志总量 时期指标 时点指标

2. 劳动单位

3. 有名数 无名数 有名数 无名数

4. 水平法 累计法

5. 比较相对数 结构相对数

6. 比较

7. 时期

8. 结构

9. 强度相对 平均

10. 可比

五、1. 销售利润率计划规定 8%，实际为 12%，超额完成计划应为 3.7%；劳动生产率计划规定比去年提高 5%，实际比去年提高 5.5%，劳动生产率完成计划应为 100.5%；产品单位成本计划规定比去年下降 3%，实际比去年下降 2.5%，实际比计划应为少下降 0.5 个百分点。

2.

项 目	2015 年		2014 年实际完成（亿元）	2015 年比 2014 年增长（%）
	实际完成（亿元）	比重（%）		
地区生产总值		100	1 826.09	
其中：第一产业	157.5			12.5
第二产业	1 080	51.43		
第三产业	862.5	41.07	789.09	8.51

3.

序号	指标名称	时期 指标	时点 指标	计划完 成程度	结构相 对指标	比例相 对指标	比较相 对指标	动态相 对指标	强度相 对指标
1	进出口总额	√							
2	在校学生数		√						
3	人均国内生产总值								√
4	我国 2014 年全年国内游客 36.1 亿人次，比上年增长 10.7%							√	
5	2014 年按实际汇率计中国 GDP 总量相当于美国的 61.7%						√		
6	人口性别比例					√			
7	升学率				√				
8	实际销售量是计划销售量的 198%			√					
9	出生人口数	√							
10	农民人均纯收入								√

续表

序号	指标名称	时期 指标	时点 指标	计划完 成程度	结构相 对指标	比例相 对指标	比较相 对指标	动态相 对指标	强度相 对指标
11	2014 年按购买力平价计, 中国 GDP 总量达 19.230 万亿美元, 高于美国 0.943 万亿美元	√							
12	2014 年我国移动电话普及率 94.5 部/百人								√
13	全市医保定点零售药店数		√						
14	商品房中销售给个人的占 93%				√				
15	2014 年全国普通高等教育本专科招生 721.40 万人, 比上年增长 3.08%							√	
16	2014 年国家全员劳动生产率为 72 313 元/人								√
17	2014 年全国招收硕士研究生 62.13 万人	√							
18	恩格尔系数				√				
19	从业人员数		√						
20	每百人拥有的私人轿车数								√

4. 实际产值 2015 年比 2014 年增长 10%; 计划规定比上年增加 8.5%;
实际产值 2015 年比 2014 年的增长率= $110\% \times (1+8.5\%) - 1 = 19.35\%$
5. 计划规定比上年增加的百分比为 $(1+6\%)/103\% - 1 = 2.9\%$;
降低成本计划完成情况指标= $1\,365/(1\,398-20) = 99.06\%$, 完成了计划, 超了 0.94%, 多降低了 13 元。
6. 略。

第 5 章

- 一、1. × 2. × 3. × 4. √ 5. × 6. × 7. × 8. √ 9. × 10. √
二、1. B 2. A 3. A 4. C 5. A 6. A 7. A 8. A 9. C 10. C
三、1. BDE 2. BC 3. ABCD 4. ACDE 5. CD 6. ABCE 7. ABC 8. CD 9. AB

10. ABCD

- 四、1. 一般水平 (代表水平 典型水平)
2. 次数 次数比重
3. 75%
4. 93.97%
5. 平均数
6. 最多 当中 位置
7. 总体标志总量 同一总体单位总量
8. 极端变量值 没有明显集中趋势
9. 11.76%
10. 集中 离散
五、1. 营业收入的平均数是 306 639.02 百万美元; 中位数是 268 566.6 百万美元。
2.

计划完成程度%	组中值 x	计划任务数 f	实际完成数 xf
90~100	0.95	80	76
100~110	1.05	400	420
110~120	1.15	120	138
合 计	—	600	634

平均计划完成程度=634/600=105.67%

3.

蔬菜名称	价格（元/千克） x	成交额（元） m	成交量（千克） m/x
土豆	2.4	2 400	1 000
油菜	3.2	6 400	2 000
蘑菇	4.4	4 400	1 000
合计	—	13 200	4 000

蔬菜的平均价格=13 200/4 000=3.3（元/千克）

4.（1）平均成绩=79分；（2）平均成绩=78.33分；（3）平均成绩=79.67分；

（4）全体学生平均成绩的变化原因在于男女生人数所占比重的变动。

5.（1）听众年龄的平均数=5 680/170=33.41岁；众数为=34.47岁。

（2）听众年龄的标准差=7.62岁

6. 中国互联网用户平均年龄=29.41岁；标准差=12.9岁。

7.（1）如果给每个标志值都加上5分钟，平均数=12.3+5=17.3分钟；

（2）如果给每个标志值都加上5分钟，标准差=4分钟；

（3）如果给每个标志值都加上5分钟，方差=16；

（4）如果给每个标志值都加上5分钟，众数、中位数、四分位数都相应增加5分钟。

8. 采用平均指标和标志变异指标评价组装方法的好坏。

	平均组装个数（个）	标准差	标准差系数%
第一种方法	128.73	1.69	1.31
第二种方法	165.6	2.06	1.24
第三种方法	165.4	6.76	4.09

因为第二种方法平均组装个数最多，标准差系数最小，所以应选第二种方法。

9. 幼儿组平均身高=71.3cm，标准差=2.37cm，标准差系数=3.32%；

成年组平均身高=172.1cm，标准差=3.99cm，标准差系数=2.32%；幼儿组的身高差异大。

10. 平均数=48.26%，中位数=45.7%，标准差系数=0.24

11. Excel函数计算结果：平均数 123.42、众数 123、中位数 123、极差 44、平均差 8.516 8、标准差 10.650 46。

利用分组资料计算结果：平均数 123.4、众数 123.571 4、中位数 123.333 3、极差 50、平均差 8.784、标准差 11.019 98。

第6章

一、1. × 2. × 3. √ 4. × 5. × 6. √ 7. × 8. × 9. √ 10. √

二、1. B 2. A 3. B 4. C 5. C 6. B 7. A 8. B 9. B 10. C

三、1. ACD 2. ABCD 3. ABCE 4. ABCE 5. ADE 6. ABCDE 7. BC 8. ACD

9. ACDE 10. BCD

四、1. 抽样调查 总体相应数量

2. 总体标准差 样本单位数

3. 不重复抽样

4. 分类抽样

5. 抽样极限误差

6. 抽样平均误差

7. 样本方差 样本成数

8. 原假设 备择假设

9. 30 个

10. 抽样极限误差 抽样平均误差

五、1. 重复抽样条件下: $\mu_p = \sqrt{\frac{P(1-P)}{n}} = 3\%$;

不重复抽样条件下: $\mu_p = \sqrt{\frac{P(1-P)}{n} \left(1 - \frac{n}{N}\right)} = 2.98\%$ 。

2. $n=400, P=0.721, t=1.96$

$$\mu_p = \sqrt{\frac{P(1-P)}{n}} = 2.24\%$$

下限= $0.721-1.96 \times 2.24\%=67.4\%$; 上限= $0.721+1.96 \times 2.24\%=76.8\%$

即我们能以 95% 的概率保证该新家电产品的市场占有率大概在 67.4%~76.8%。

3. (1) 抽样极限误差=500

$$t=1.96; \text{ 样本容量 } n = \frac{t^2 \sigma^2}{\Delta_{\bar{x}}^2} = 27.66 \approx 28$$

(2) 由于 $n = \frac{t^2 \sigma^2}{\Delta_{\bar{x}}^2}$, 其中 $n=56$, 则抽样极限误差=351.4, 所以估计值处在总体平均数附近 351.4 元的

范围内。

4. $n=200, P=0.04$

$$\mu_p = \sqrt{\frac{P(1-P)}{n} \left(1 - \frac{n}{N}\right)} = 1.35\%$$

下限= $0.04-2 \times 1.35\%=1.3\%$; 上限= $0.04+2 \times 1.35\%=5.3\%$

所以, 不能认为这批产品的废品率不超过 5%。

$$5. \mu=176; H_1: \mu \neq 176, t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = 2.2280, t(16) = \pm 2.1199,$$

由于 $t=2.2280 > 2.1199$, 因此拒绝 H_0 , 即在 0.05 显著性水平下, 可以推断美国总统的身高与普通男人身高有显著差异。

6. 及格率区间为 74.32%~89.68%。

在其他条件不变时, 允许误差缩小一半, 应抽取 400 名学生。

7. (1) 样本平均数=150.3; 样本方差=0.76; 抽样平均误差=0.087;

当 $t=3$ 时, 下限= $150.3-3 \times 0.087=150.04$; 上限= $150.3+3 \times 0.087=150.56$

即 150.04~150.56 克; 表明这批食品平均每包重量达到规格要求。

(2) 样本合格率: 70%

$$\text{抽样平均误差: } \mu_p = \sqrt{\frac{P(1-P)}{n}} = 0.0458$$

当 $t=3$ 时, 下限= $0.7-3 \times 0.0458=56.26\%$; 上限= $0.7+3 \times 0.0458=83.74\%$

即这批食品包装的合格率在 56.26%~83.74%。

8. 因为置信区间宽度为 4 个百分点, 所以抽样极限误差为 2%。

$$n = \frac{t^2 P(1-P)}{\Delta_p^2} = 1\,536.64 \approx 1\,537$$

$$9. N=2\,000, t=2, P=60\%$$

$$\mu_p = \sqrt{\frac{P(1-P)}{n} \left(1 - \frac{n}{N}\right)} = 4.77\%;$$

$$\text{下限} = 0.60 - 2 \times 4.77\% = 50.46\%; \quad \text{上限} = 0.60 + 2 \times 4.77\% = 69.54\%$$

以 95.45% 的概率估计该校学生英语等级考试成绩在 70 分以上的学生所占比重范围为 50.46%~69.54%。

$$10. \text{ 抽样平均误差: } \mu_p = \sqrt{\frac{P(1-P)}{n}} = 4.05\%;$$

以 95% 把握程度推断出小区中赞成该项改革的户数比例的置信区间为 56.35%~72.23%。

第 7 章

一、1. × 2. √ 3. × 4. × 5. √ 6. × 7. √ 8. √ 9. × 10. √

二、1. D 2. A 3. C 4. C 5. C 6. B 7. A 8. D 9. B 10. C

三、1. CE 2. ABC 3. BC 4. ABD 5. ABC 6. ABCD 7. AD 8. BDE 9. BDE 10. ACD

四、1. 时期 数值

2. 时期数列 时点数列

3. 时点 时期

4. 现象的水平分析 速度分析

5. 变量数列 动态数列

6. 33.06%

7. 连乘积 和

8. 相应的环比发展速度 前一期的累计增长量

9. 0.94%

10. 按月（季）平均法 移动平均趋势剔除法

五、1. 300.3 辆

2. 2009—2013 年我国餐饮业平均从业人数为 228.875 万人。

3. 84.4%

4. (1) 197.3 万元 (2) 205.3 人 (3) 2.88 万元

(4) 4 月份的=0.98 万元 5 月份的=0.86 万元 6 月份的=1.05 万元

(5) 0.96 万元

5. (1)

年 份	产量（万台）	累计增长量（万台）	定基发展速度（%）	环比发展速度（%）
2010		0	100	—
2011	250		125	125
2012	250	50		100
2013	287.5	87.5	143.75	
2014	359.4	159.4	179.7	
2015	360		180	100.2

(2) 平均增长量=32 万台 (3) 平均发展速度=112.5% (4) 平均增长速度=12.5%

6. 2020 年将达到水平=636 463×1.06⁶=902 834.9（亿元）。

7. 如果以后每年增长 6.5%，需要 11 年后才能达到 200 亿元。

8. (1) ~ (4) 小题计算结果见下表。

年 份	2009	2010	2011	2012	2013	2014	2015
销售额 (百万元)	5.6	11.4	17.1	24	31.3	38	43.9
逐期增长量 (百万元)	—	5.8	5.7	6.9	7.3	6.7	5.9
累计增长量 (百万元)	0	5.8	11.5	17.4	24.7	31.4	37.3
定基发展速度	100	203.57	305.36	428.57	558.93	678.57	783.93
环比发展速度	—	203.57	150	140.35	130.42	121.41	115.53
增长 1% 的绝对值	—	0.056	0.114	0.171	0.24	0.313	0.38

(5) 平均增长速度=140.94%

(6)

年 份 t	-3	-2	-1	0	1	2	3
销售额 (百万元) y	5.6	11.4	17.1	24	31.3	38	43.9

$$\begin{cases} a = \frac{\sum y}{n} \\ b = \frac{\sum ty}{\sum t^2} \end{cases}$$

$b=6.5107; a=24.47$

$y_c=24.47+6.5107t$

当 $t=4$ 时, 2016 年销售额=50.51 百万元

9. (1) 季节变动 (2) 季节变动 (3) 长期趋势 (4) 不规则变动 (5) 循环变动

(6) 季节变动 (7) 季节变动 (8) 季节变动

10. (1) 四项移动平均并经二次移动后数列为

单位: 万吨

年 份	一季	二季	三季	四季
2012			13.125	13.375
2013	13.75	14.25	13.5	13.75
2014	15.25	16	16.875	18.125
2015	19.625	20.625		

(2)

年 份	一季	二季	三季	四季	合 计	季平均数
2012	15	7	10	20	52	13
2013	16	8	12	20	56	14
2014	18	10	14	24	66	16.5
2015	21	17	19	27	84	21
合计	70	42	55	91	258	64.5
同季平均数	17.5	10.5	13.75	22.75	64.5	16.125
季节比率%	108.53	65.12	85.27	141.08	400	100

(3) 假设 2016 年全年计划收购量为 120 万吨, 则将计划目标分解, 一、二、三、四季度计划收购量应分别为 32.559、19.536、25.581、42.324 万吨。

11.略。

12.（1）按季平均法测定季节比率：

年 份	第 1 季度	第 2 季度	第 3 季度	第 4 季度	合 计
2012	450	150	250	500	1 350
2013	500	180	280	560	1 520
2014	580	200	320	630	1 730
2015	700	250	380	750	2 080
合计	2 230	780	1 230	2 440	6 680
同季平均	557.5	195	307.5	610	417.5
季节比率%	133.5	46.7	73.64	146.07	399.91
调整季节比率%	133.53	46.71	73.66	146.1	400

（2）移动平均趋势剔除法测定季节比率：

年 份	季 度	销售量（Y）	一次移动	移正平均 T	Y/T*100	四舍五入的结果
2012	1	450	—	—	—	—
	2	150	—	—	—	—
	3	250	337.5	343.75	72.727 273	72.73
	4	500	350	353.75	141.342 76	141.34
2013	1	500	357.5	361.25	138.408 3	138.41
	2	180	365	372.5	48.322 148	48.32
	3	280	380	390	71.794 872	71.79
	4	560	400	402.5	139.130 43	139.13
2014	1	580	405	410	141.463 41	141.46
	2	200	415	423.75	47.197 64	47.2
	3	320	432.5	447.5	71.508 38	71.51
	4	630	462.5	468.75	134.4	134.4
2015	1	700	475	482.5	145.077 72	145.08
	2	250	490	505	49.504 95	49.5
	3	380	520	—	—	—
	4	750	—	—	—	—

利用移动平均趋势剔除法剔除长期趋势后季节比率计算如下表：

年 份	羊绒衫销售量				合 计
	第 1 季度	第 2 季度	第 3 季度	第 4 季度	
2012	—	—	72.73	141.34	1 201.11
2013	138.41	48.32	71.79	139.13	
2014	141.46	47.2	71.75	134.4	
2015	145.08	49.5	—	—	
合计	424.95	145.02	216.27	414.87	
同季平均	141.65	48.34	72.09	138.29	100.092 5
季节比率%	141.519 1	48.295 33	72.023 38	138.162 2	400

13.略。

第 8 章

- 一、1. × 2. × 3. × 4. × 5. × 6. × 7. × 8. × 9. × 10. √
- 二、1. B 2. D 3. C 4. B 5. B 6. A 7. B 8. C 9. C 10. C
- 三、1. ABC 2. CE 3. BCE 4. ACD 5. AE 6. AB 7. ACD 8. ADE 9. ABD 10. CDE
- 四、1. 加权算术平均指数 加权调和平均指数
2. 降低了 9.1%
3. 6.05%
4. 指数
5. 质量指标 基 数量指标 报告
6. 个体指数 算术 调和
7. 假设在报告期产量的情况下, 报告期成本比基期降低了 3% 由于单位成本的降低使总成本报告期比基期减少了 2 200 元
8. 105.26%
9. 下降 10%
10. 上升 1.03%
- 五、1.

统计指数	个体指数	总指数	数量指标指数	质量指标指数
某一种商品单位成本指数	√			
二十种商品的价格指数		√		√
企业职工的平均工资指数		√		√
某型号手机的销售量指数	√			
全国消费品零售价格指数		√		√

2. (1) 甲: 114.29% 105%
乙: 100% 100%
丙: 97% 110%
- (2) 三种商品的物价总指数=100.79% 三种商品的销售量总指数=105.83%
- (3) 三种商品的销售额总指数=106.67%
- (4) 106.67%=105.83%×100.79% 400 万元=350 万元+50 万元
3. $\bar{k}_q = \frac{\sum k_q q_0 p_0}{\sum q_0 p_0} = 98.64\%$, 17 607-17 850=-243 (元)
- 4 月与 3 月相比销售量平均 (总共) 下跌 1.36%, 销售额减少 243 元。
4. $\bar{k}_p = \frac{\sum p_1 q_1}{\sum \frac{1}{k_p} p_1 q_1} = 118.36\%$, 23 600-19 939.4=3 660.6 (元)
- 4 月与 3 月相比销售价格平均 (总共) 上升 18.36%, 销售额增加了 3 660.6 元。
5. 单价的变动对产值的变动影响 114.77%; 169.88 万元
产品产量的变动对产值的变动影响 106.07%; 80.12 万元
6. (1) 价格总指数=99.42% 销售量总指数=129.55%
- (2) 在总销售额增长绝对值中, 受价格因素影响的绝对值=-200 万元
在总销售额增长绝对值中, 受销售量因素影响的绝对值=7 800 万元
7 600 万元=(-200 万元)+7 800 万元

7. (1) 生产总成本指数=107.5% (2) 产品产量指数=110.8%
 (3) 由于单位成本降低而节约的绝对额=-0.396 万元
 8. (1) 102% 2015 年与 2014 年比较, 零售物价上升 2%; 绝对量=49.3 亿元
 (2) 零售量指数=107.30%; 绝对量=171.5 亿元
 109.4%=107.30%×102% 220.8 亿元=171.5 亿元+49.3 亿元
 9. (1) 2014 年的总平均成本=9.96 元; 2015 年的总平均成本=9.23 元
 (2) 平均成本指数=92.7%
 (3) 92.7%=109.39%×84.72% -0.73=0.792+(-1.522)
 10. (1) 价格指数为: 88.695 7%; 销售量指数为: 111.854 1%
 (2) 销售额指数=价格指数×销售量指数; 即 99.209 7%=88.695 7%×111.854 1%
 由于价格变动对销售量的影响数为: -2 080
 由于销售量变动对销售额的影响数为: 1 950
 销售额总变动为: -130
 销售额总变动=由于价格变动对销售量的影响数+由于销售量变动对销售额的影响数
 即-130=-2 080+1 950

第9章

- 一、1. × 2. × 3. √ 4. √ 5. √ 6. √ 7. √ 8. × 9. √ 10. √
 二、1. C 2. C 3. C 4. D 5. C 6. C 7. B 8. B 9. B 10. D
 三、1. ABCD 2. AB 3. DE 4. BD 5. ABD 6. BCE 7. ABD 8. BCDE
 9. ADE 10. DE
 四、1. 函数 -1 或+1
 2. 相关
 3. 估计标准误差
 4. 方程的起点值 回归系数 x 每增加一个单位 y 的平均增加值 最小平方
 5. 前提和基础 深入和继续
 6. 平均数 回归方程
 7. 完全负相关 完全正相关 完全无直线相关
 8. 0 到 1
 9. 代表性最差 代表性最好
 10. $0 < |r| < 0.3$ 为微相关 $0.3 \leq |r| < 0.5$ 为低度相关 $0.5 \leq |r| < 0.8$ 为显著相关 $0.8 \leq |r| < 1$ 为高度相关
 五、1. (1) 编制相关表如下:

航班正点率 (%)	91.4	81.8	76.6	76.6	75.7	73.8	72.2	71.2	70.8	68.5
顾客投诉次数	18	21	58	85	68	74	93	72	122	125

相关图略。

由此可见, 航班正点率和顾客投诉次数之间成负相关关系。

- (2) 相关系数=-0.868 6
 (3) 回归方程 $y_c = 430.19 - 4.700 6x$ 回归系数的意义是航班正点率每提高 1%, 顾客投诉次数平均下降 4.700 6 次。
 (4) 如果航班正点率为 80%, 顾客的投诉次数=54。
 2. (1) 用广告费支出作自变量, 销售量作因变量, 估计的回归方程 $y_c = 4.07 + 0.195 8x$

- (2) 广告费用每增加 10 000 元, 销售量将上升 1 958 箱。
3. (1) 飞行成本 y 随乘客数量 x 的直线回归方程 $y_c = 1.57 + 0.0407x$ 。
(2) 飞行每增加一名乘客, 飞行成本将增加 40.7 美元。
(3) 在此案例中, 如果飞行时不搭载乘客, 飞行 500 英里的飞行成本为 1 570 美元。
4. (1) 账单数额和小费数额之间相关系数=0.828 2, 所以两者之间存在高度正相关关系。
(2) 账单数额和小费数额之间回归方程为 $y_c = -0.35 + 0.1486x$
当账单数额为 100 美元时, 应该留下小费数=14.51 美元。
(3) 现已知给出小费 10 元, 回归方程为 $y_c = 24.74 + 4.6150x$
其消费的账单数额为 70.89 美元。
5. (1) 相关关系, 画散点图
(2) 相关系数=0.851 8
(3) 直线回归方程 $y_c = 117 + 9.74x$
(4) 气温每升高 1°C , 销售量平均增加 9.74 箱。
6. 相关系数=0.144 2, 优秀的学生学习时间与平均成绩之间相关程度低。
7. 略。
8. 略。
9. (1) 相关系数为: 0.993 402; (2) 回归方程为: $y_c = -7.273 + 0.074x$
10. (1) 相关系数为: 0.749; (2) 回归方程为: $y_c = 0.000413 + 0.05883x$

附录 B 正态分布概率表

t	$F(t)$	t	$F(t)$	t	$F(t)$	t	$F(t)$
0.00	0.000 0	0.32	0.251 0	0.64	0.477 8	0.96	0.662 9
0.01	0.008 0	0.33	0.258 6	0.65	0.484 3	0.97	0.668 0
0.02	0.016 0	0.34	0.266 1	0.66	0.490 7	0.98	0.672 9
0.03	0.023 9	0.35	0.273 7	0.67	0.497 1	0.99	0.677 8
0.04	0.031 9	0.36	0.281 2	0.68	0.503 5	1.00	0.682 7
0.05	0.033 9	0.37	0.288 6	0.69	0.509 8	1.01	0.687 5
0.06	0.047 8	0.38	0.296 1	0.70	0.516 1	1.02	0.692 3
0.07	0.055 8	0.39	0.303 5	0.71	0.522 3	1.03	0.697 0
0.08	0.063 8	0.40	0.310 8	0.72	0.528 5	1.04	0.701 7
0.09	0.071 7	0.41	0.318 2	0.73	0.534 6	1.05	0.706 3
0.10	0.079 7	0.42	0.325 5	0.74	0.540 7	1.06	0.710 9
0.11	0.087 6	0.43	0.332 8	0.75	0.546 7	1.07	0.715 4
0.12	0.095 5	0.44	0.340 1	0.76	0.552 7	1.08	0.719 9
0.13	0.103 4	0.45	0.347 3	0.77	0.558 7	1.09	0.724 3
0.14	0.103 4	0.46	0.354 5	0.78	0.564 6	1.10	0.728 7
0.15	0.119 2	0.47	0.361 6	0.79	0.570 5	1.11	0.733 0
0.16	0.127 1	0.48	0.368 8	0.80	0.576 3	1.12	0.737 3
0.17	0.135 0	0.49	0.375 9	0.81	0.582 1	1.13	0.741 5
0.18	0.142 8	0.50	0.382 9	0.82	0.587 8	1.14	0.745 7
0.19	0.150 7	0.51	0.389 9	0.83	0.593 5	1.15	0.749 9
0.20	0.158 5	0.52	0.396 9	0.84	0.599 1	1.16	0.754 0
0.21	0.166 3	0.53	0.403 9	0.85	0.604 7	1.17	0.758 0
0.22	0.174 1	0.54	0.410 8	0.86	0.610 2	1.18	0.762 0
0.23	0.181 9	0.55	0.417 7	0.87	0.615 7	1.19	0.766 0
0.24	0.189 7	0.56	0.424 5	0.88	0.621 1	1.20	0.769 9
0.25	0.197 4	0.57	0.431 3	0.89	0.626 5	1.21	0.773 7
0.26	0.205 1	0.58	0.438 1	0.90	0.631 9	1.22	0.777 5
0.27	0.212 8	0.59	0.444 8	0.91	0.637 2	1.23	0.781 3
0.28	0.220 5	0.60	0.451 5	0.92	0.642 4	1.24	0.785 0
0.29	0.228 2	0.61	0.458 1	0.93	0.647 6	1.25	0.788 7
0.30	0.235 8	0.62	0.464 7	0.94	0.652 8	1.26	0.792 3
0.31	0.243 4	0.63	0.473 1	0.95	0.657 9	1.27	0.795 9

续表

t	$F(t)$	t	$F(t)$	t	$F(t)$	t	$F(t)$
1.28	0.799 5	1.61	0.892 6	1.94	0.947 6	2.54	0.795 9
1.29	0.803 0	1.62	0.894 8	1.95	0.948 8	2.56	0.988 9
1.30	0.806 4	1.63	0.896 4	1.96	0.950 0	2.58	0.989 5
1.31	0.809 8	1.64	0.899 0	1.97	0.951 2	2.60	0.990 1
1.32	0.813 2	1.65	0.901 1	1.98	0.952 3	2.62	0.990 7
1.33	0.816 5	1.66	0.903 1	1.99	0.953 4	2.64	0.991 2
1.34	0.819 8	1.67	0.905 1	2.00	0.954 5	2.66	0.991 7
1.35	0.823 0	1.68	0.907 0	2.02	0.956 6	2.68	0.992 2
1.36	0.826 2	1.69	0.909 0	2.04	0.958 7	2.70	0.992 6
1.37	0.829 3	1.70	0.910 9	2.06	0.960 6	2.72	0.993 1
1.38	0.832 4	1.71	0.912 7	2.08	0.962 5	2.74	0.993 5
1.39	0.835 5	1.72	0.914 6	2.10	0.964 3	2.76	0.993 9
1.40	0.838 5	1.73	0.916 4	2.12	0.966 0	2.78	0.994 2
1.41	0.841 5	1.74	0.918 1	2.14	0.967 6	2.80	0.994 6
1.42	0.844 4	1.75	0.919 9	2.16	0.969 2	2.82	0.994 9
1.43	0.847 3	1.76	0.921 6	2.18	0.970 7	2.84	0.995 2
1.44	0.850 1	1.77	0.923 3	2.20	0.972 2	2.86	0.995 5
1.45	0.852 9	1.78	0.924 9	2.22	0.973 6	2.88	0.995 8
1.46	0.855 7	1.79	0.926 5	2.24	0.974 9	2.90	0.996 0
1.47	0.858 4	1.80	0.928 1	2.26	0.976 2	2.92	0.996 2
1.48	0.861 1	1.81	0.929 7	2.28	0.977 4	2.94	0.996 5
1.49	0.863 8	1.82	0.931 2	2.30	0.978 6	2.96	0.996 7
1.50	0.866 4	1.83	0.932 8	2.32	0.979 7	2.98	0.996 9
1.51	0.869 0	1.84	0.934 2	2.34	0.980 7	3.00	0.997 3
1.52	0.871 5	1.85	0.935 7	2.36	0.981 7	3.20	0.997 9
1.53	0.874 0	1.86	0.937 1	2.38	0.982 7	3.40	0.998 6
1.54	0.876 4	1.87	0.938 5	2.40	0.983 6	3.80	0.999 68
1.55	0.878 9	1.88	0.939 9	2.42	0.984 5	4.00	0.999 94
1.56	0.881 2	1.89	0.941 2	2.44	0.985 3	4.50	0.999 993
1.57	0.883 6	1.90	0.942 6	2.46	0.986 1	5.00	0.999 999
1.58	0.885 9	1.91	0.943 9	2.48	0.986 9		
1.59	0.888 2	1.92	0.945 1	2.50	0.987 6		
1.60	0.890 4	1.93	0.946 4	2.52	0.988 3		

附录 C 随机数表（摘录）

	112345678910	212345678910	312345678910	412345678910	512345678910
1	6 119 690 446	2 645 747 774	5 192 433 729	6 539 459 593	4 258 260 527
2	1 547 445 266	9 527 079 953	5 936 783 848	8 239 610 118	3 321 159 466
3	9 455 728 573	6 789 754 387	5 462 244 431	9 119 042 592	9 292 745 973
4	4 248 116 213	9 734 408 721	1 686 848 767	0 307 112 059	2 570 146 670
5	2 352 378 317	7 320 889 837	6 893 591 416	2 625 229 663	0 552 282 562
6	0 449 352 494	7 524 633 824	4 586 251 025	6 196 279 335	6 533 712 472
7	0 054 997 654	6 405 188 159	9 611 963 896	5 469 282 391	2 328 729 529
8	3 596 315 307	2 689 809 354	3 335 135 462	7 797 450 024	9 010 339 333
9	5 980 808 391	4 542 726 842	8 360 949 700	1 302 124 892	7 856 520 106
10	4 605 885 236	0 139 092 286	7 728 144 077	9 391 083 647	7 061 742 941
11	3 217 900 597	8 737 925 241	0 556 707 007	8 674 317 157	8 539 411 838
12	6 923 461 406	2 011 745 204	1 595 660 000	1 874 392 423	9 711 896 388
13	1 956 541 430	0 175 875 379	4 041 921 585	6 667 436 806	8 496 285 207
14	4 515 514 938	1 947 607 246	4 366 794 543	5 904 790 033	2 082 669 541
15	9 486 431 994	3 616 810 851	3 488 881 553	0 154 035 456	0 501 451 176
16	9 808 624 826	4 524 028 404	4 499 908 896	3 909 473 407	3 544 131 880
17	3 318 516 232	4 194 150 949	8 943 548 581	8 869 541 904	3 754 873 043
18	8 095 100 406	9 638 270 774	2 015 123 387	2 501 625 298	9 462 461 171
19	7 975 249 140	7 196 128 296	6 986 102 591	7 485 220 539	0 038 759 579
20	1 863 332 537	9 814 506 571	3 101 024 674	0 545 561 427	7 793 891 936

附录 D t -分布临界值表

$P\{t(k)>t_{\alpha}\}=\alpha$

$t(k)$ t_{α}	0.25	0.10	0.05	0.025	0.01	0.005
1	1.000 0	3.077 7	6.313 8	12.706 2	31.820 7	63.657 4
2	0.816 5	1.885 6	2.920 0	4.320 7	6.964 6	9.924 8
3	0.764 9	1.637 7	2.353 4	3.182 4	4.540 7	5.840 9
4	0.740 7	1.533 2	2.131 8	2.776 4	3.746 9	4.604 1
5	0.726 7	1.475 9	2.015 0	2.570 6	3.364 9	4.032 2
6	0.717 6	1.439 8	1.943 2	2.446 9	3.142 7	3.707 4
7	0.711 1	1.414 9	1.894 6	2.364 6	2.998 0	3.499 5
8	0.706 4	1.396 8	1.859 5	2.306 0	2.896 5	3.355 4
9	0.702 7	1.383 0	1.833 1	2.262 2	2.821 4	3.249 8
10	0.699 8	1.372 2	1.812 5	2.228 1	2.763 8	3.169 3
11	0.697 4	1.363 4	1.795 9	2.201 0	2.718 1	3.105 8
12	0.695 5	1.356 2	1.782 3	2.178 8	2.681 0	3.054 5
13	0.693 8	1.350 2	1.770 9	2.160 4	2.650 3	3.012 3
14	0.692 4	1.345 0	1.761 3	2.144 8	2.624 5	2.976 8
15	0.691 2	1.340 6	1.753 1	2.131 5	2.602 5	2.946 7
16	0.690 1	1.336 8	1.745 9	2.119 9	2.583 5	2.902 8
17	0.689 2	1.333 4	1.739 6	2.109 8	2.566 9	2.898 2
18	0.688 4	1.330 4	1.734 1	2.100 9	2.552 4	2.878 4
19	0.687 6	1.327 7	1.729 1	2.093 0	2.539 5	2.860 9
20	0.687 0	1.325 3	1.724 7	2.086 0	2.528 0	2.845 3
21	0.684 4	1.323 2	1.720 7	2.079 6	2.517 7	2.831 4
22	0.685 8	1.321 2	1.717 1	2.073 9	2.508 3	2.818 8
23	0.685 3	1.319 5	1.713 9	2.068 7	2.499 9	2.807 3
24	0.684 8	1.317 8	1.710 9	2.063 9	2.492 2	2.796 9
25	0.684 4	1.316 3	1.708 1	2.059 5	2.485 1	2.787 4

续表

t_{α} $t(k)$	0.25	0.10	0.05	0.025	0.01	0.005
26	0.684 0	1.315 0	1.705 6	2.055 5	2.478 6	2.778 7
27	0.683 7	1.313 7	1.703 3	2.051 8	2.472 7	2.770 7
28	0.683 4	1.312 5	1.701 1	2.048 4	2.467 1	2.763 3
29	0.683 0	1.311 4	1.699 1	2.045 2	2.462 0	2.756 4
30	0.682 8	1.310 4	1.697 3	2.042 3	2.457 3	2.750 0

参考文献

- [1] 阮红伟. 统计学. 第2版. 北京: 北京大学出版社, 2014.
- [2] 李洁明, 祁新娥. 统计学原理. 第6版. 上海: 复旦大学出版社, 2014.
- [3] [美]查尔斯·惠伦著. 赤裸裸的统计学. 曹槟译. 北京: 中信出版社, 2013.
- [4] [美]尼尔·J·萨尔金德著. 爱上统计学. 第2版. 史玲玲译. 重庆: 重庆大学出版社, 2011.
- [5] Ken Black. Business Statistics:For Contemporary Decision Making. Hongkong:Leyh Publishing, LLC,2004.
- [6] [美]詹姆斯·R·埃文斯著. 商业统计学精要. 潘文卿, 丁海山译. 北京: 中国人民大学出版社, 2004.
- [7] [美]Mario F.Triola 著. 初级统计学. 第8版. 刘新立译. 北京: 清华大学出版社, 2004.
- [8] [美]戴维·S·穆尔著. 统计学的世界. 第5版. 郑维厚译. 北京: 中信出版社, 2003.
- [9] [美]戴维 R.安德森著. 商务与经济统计. 第9版. 张建华等译. 北京: 机械工业出版社, 2006.
- [10] [美]蒂莫西·C·厄丹著. 白话统计学. 第3版. 彭志文译. 北京: 中国人民大学出版社, 2013.

反侵权盗版声明

电子工业出版社依法对本作品享有专有出版权。任何未经权利人书面许可，复制、销售或通过信息网络传播本作品的行为；歪曲、篡改、剽窃本作品的行为，均违反《中华人民共和国著作权法》，其行为人应承担相应的民事责任和行政责任，构成犯罪的，将被依法追究刑事责任。

为了维护市场秩序，保护权利人的合法权益，我社将依法查处和打击侵权盗版的单位和个人。欢迎社会各界人士积极举报侵权盗版行为，本社将奖励举报有功人员，并保证举报人的信息不被泄露。

举报电话：(010) 88254396; (010) 88258888

传 真：(010) 88254397

E - m a i l: dbqq@phei.com.cn

通信地址：北京市万寿路 173 信箱

电子工业出版社总编办公室

邮 编：100036